Report No. SFIM-AEC-ET-CR-95033

U.S. Army
Environmental
Center

Evaluation of Individual Demonstrator Performance at the Unexploded Ordnance Advanced Technology Demonstration Program at Jefferson Proving Ground (Phase I)

March 1995

Prepared by PRC Inc.

Distribution Unlimited; Approved for Public Release

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave Blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	30 March 1995	August 1993 - December 1994

4. TITLE AND SUBTITLE

Evaluation of Individual Demonstrator Performance at the Unexploded Ordnance Advanced Technology Demonstration Program at Jefferson Proving Ground (Phase I)

5. FUNDING NUMBERS

6. AUTHOR(S)

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Naval Explosive Ordnance Disposal Technology Division
Project Manager: Gerard Snyder 301/743-6855
Senior Engineer: M. Andrew Pedersen 301/743-6854
2008 Stump Neck Road
Indian Head, Maryland 20640-5070

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Environmental Center
Project Officer: Kelly Rigano 410/612-6868
SFIM-AEC-ETP
Aberdeen Proving Ground, Maryland 21010-5401

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

SFIM-AEC-ET-CR-95033

11. SUPPLEMENTARY NOTES

Supporting Contractor: Institute for Defense Analyses
1801 N. Beauregard Street
Alexandria, Virginia 22311-1772

12a. DISTRIBUTION / AVAILABILITY STATEMENT

Unlimited Distribution

This document has been approved for public release and sale; its distribution is unlimited.

12b. DISTRIBUTION CODE

"A"

13. ABSTRACT (Maximum 200 words)

The data contained in this report is a supplement to report SFIM-AEC-ET-CR-94120, "Unexploded Ordnance Advanced Technology Demonstration Program at Jefferson Proving Ground (Phase I)". This report provides a further analysis of the individual demonstrators and the performance of their systems when used to detect, identify and/or remediate buried unexploded ordnance under realistic, controlled conditions.

14. SUBJECT TERMS

Unexploded Ordnance, Detection, Identification, Remediation

15. NUMBER OF PAGES

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

NSN 7540-01-280-5500

Standard Form 298 (Rev 12-94)
Prescribed by ANSI Sta 239-18
298-102

EVALUATION OF INDIVIDUAL DEMONSTRATOR PERFORMANCE AT THE UNEXPLODED ORDNANCE ADVANCED TECHNOLOGY DEMONSTRATION PROGRAM AT JEFFERSON PROVING GROUND (PHASE 1)

MARCH 1995

This page is intentionally left blank

CONTENTS

FIGURES

TABLE

This page is intentionally left blank

1. INTRODUCTION

It is inevitable that some ordnance does not explode as intended. Unexploded ordnance (UXO) remains as the legacy of past testing, training, and wartime activities. Millions of acres of government-owned or previously government-owned land in the United States and in foreign countries is contaminated with unexploded ordnance. In some cases the ordnance is over one hundred years old, and in other cases it is sophisticated modern ordnance. The ordnance may be on the surface and hence visible, or it may be completely buried. UXO contamination is found on active ranges, on bases slated for realignment and closure, on formerly used defense sites (FUDS), on other government lands such as Department of Interior or Fish and Wildlife lands. The United States Department of Defense (DoD) has the responsibility for remediation of domestic land contaminated with UXO.

Current methods for clearing unexploded ordnance from contaminated land are labor-intensive, hazardous, and costly. Furthermore, a great deal of controversy exists concerning the capabilities of systems in current use or proposed for use. To address this issue, and to encourage the timely development and demonstration of technology to detect and remediate UXO contamination, the United States Congress mandated funds for a UXO technology demonstration. Subsequently, the Army Environmental Center (AEC), as program manager, and the Naval Explosive Ordnance Disposal Technology Division (NAVEODTECHDIV), for technical support, conducted a large-scale UXO Technology Demonstration at the Jefferson Proving Ground (JPG), Indiana.

The report entitled *Unexploded Ordnance Advanced Technology Demonstration Program at Jefferson Proving Ground (Phase 1) 23 Dec 94,* provides an overview of the results from the demonstration, as well as a description of the demonstrations sites, their preparation, and the demonstrator systems. The information contained in this second report presents a more detailed

analysis of the performance of systems demonstrated at JPG. Also presented are alternative evaluation techniques and corrections to the data that facilitate additional comparisons among demonstrators. As such, this report is intended for those interested in the detailed performance of individual demonstrators.

2. DESCRIPTION OF EVALUATION CRITERIA FOR DEMONSTRATORS OF DETECTION SYSTEMS

Demonstrators of UXO detection technologies have been evaluated using the following criteria (Ref. 1), listed in decreasing order of importance:

1. Detection Capability
2. False Negative Rate
3. False Positive Rate
4. Target Position and Accuracy
5. Target Classification Capability
6. Survey Rate
7. Survey Costs

A discussion of the method used to measure each of these seven criteria follows.

Before quantitative measures of the scoring criteria can be generated, the demonstrator declarations must be appropriately matched to the emplaced baseline items to determine which declarations will be "detections" and which will be "false alarms." Because both ordnance and nonordnance items were emplaced, and because the demonstrators typed their declarations as either ordnance or nonordnance, there are several categories of both detections and false alarms. A Venn diagram is useful in describing the different sets of ordnance and nonordnance targets and demonstrator declarations. The circle on the right in the Venn diagram shown in Fig. 1 contains all of the emplaced objects (*B*). This set is referred to as the baseline, and may be divided into ordnance (*BO*) and nonordnance (*BN*) sets. The circle on the left contains all of the targets declared by the demonstrator (*D*), which can be similarly described. The line through each circle separates ordnance targets from nonordnance targets. The variables defined in the Venn diagram and described in the following table are used to define formulas for evaluating demonstrator performance.

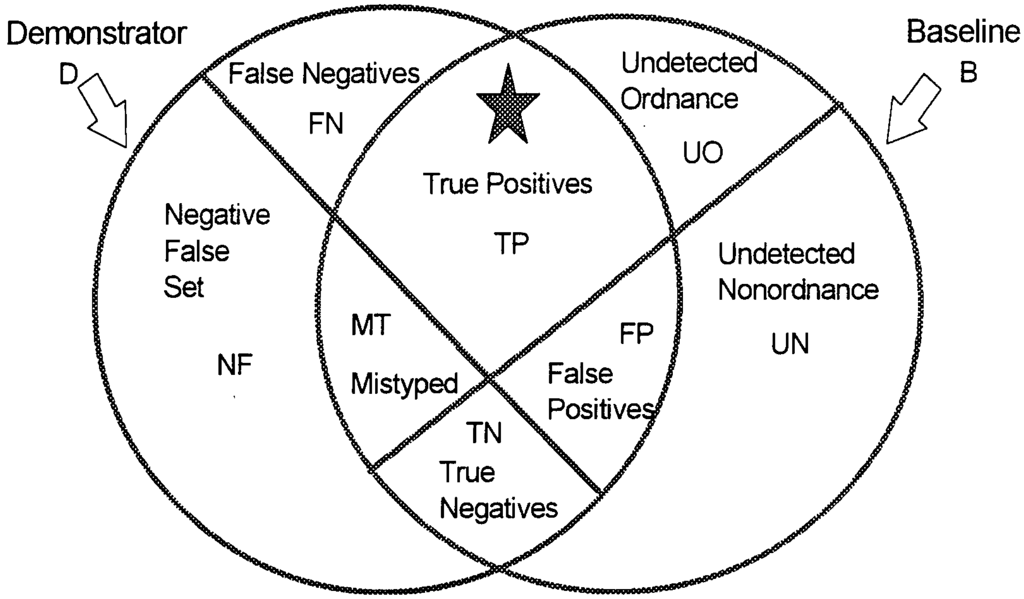


Figure 1. Venn Diagram

Variable Name	Symbol	Description
Demonstrator Ordnance Set	DO	Demonstrator ordnance declarations
Baseline Ordnance Set	BO	Baseline ordnance items
Demonstrator Nonordnance Set	DN	Demonstrator nonordnance declarations
Baseline Nonordnance Set	BN	Baseline nonordnance items
Detected Target Set	E	Demonstrator declaration determined to match an emplaced item E = TP + MT + TN + FP
True Positive Set	TP	Baseline ordnance items detected and identified as ordnance by the demonstrator
Mistyped Target Set	MT	Baseline ordnance items detected by the demonstrator but not identified as ordnance
True Negative Set	TN	Baseline nonordnance items detected and identified as nonordnance by the demonstrator
False Positive Set	FP	Baseline nonordnance items detected by the demonstrator, but identified as ordnance
False Negative Set	FN	Items that are declared by the demonstrator as ordnance but not matched to baseline items
Negative False Set	NF	Items that are declared by the demonstrator as nonordnance but not matched to baseline items
Undetected Ordnance Set	UO	Baseline ordnance items not detected by the demonstrator
Undetected Nonordnance Set	UN	Baseline nonordnance items not detected by the demonstrator

A target matching algorithm constructed by Automation Research Systems, Limited was used to associate demonstrator declarations with emplaced items. Elements of the demonstrator ordnance and nonordnance sets *DO* and *DN* are placed in the sets described by the Venn diagram depending upon their association with elements of the baseline ordnance and nonordnance sets *BO* and *BN*. In this algorithm, R_{crit} is used to define the maximum horizontal separation between a baseline item and a demonstrator declaration for the two to be further considered as a potential match. The selection of R_{crit} will influence the results; for example, the probability of a match between a baseline item and a demonstrator declaration, P_{match}, will increase monotonically with R_{crit}. $R_{crit} = 2$ m was used to determine the numbers tabulated in the demonstrator evaluations for ground based systems. For airborne systems, R_{crit} values of 2 m, 5 m, and 10 m were used.

To minimize the effect of an arbitrary choice of R_{crit} on the results, two issues were considered. First, R_{crit} was chosen at a point where the change in P_{match} with R_{crit} was slow. Second, for some demonstrators it was useful to apply a correction to P_{match} that reduces P_{match} by the probability of finding a target at random in the fraction of the site area covered by the demonstrator declarations at a given R_{crit}. This corrected P_{match} is flat for most demonstrators in the region 1 m $< R_{crit} <$ 3 m. This result gives confidence that the standard R_{crit} values used capture all the non-accidental detections. For demonstrators with very high false alarm rates, the corrected P_{match} may give a more accurate indication of performance. For some demonstrators, negative values are reported for the corrected P_{match}, indicating that more matches are expected if an equal number of declarations are placed randomly on the site. The values tabulated in the demonstrator reports do not include this correction unless so indicated.

In the following paragraphs the notation $|X|$ is used to indicate the number of items in set *X*.

Criteria for Measuring Demonstrator Performance

The measures for evaluating demonstrator performance were calculated for both the entire site and the portion of the site that the demonstrator reported visiting. One reason for measuring demonstrator performance on the entire site, whether it was visited or not, is consistency. Different areas of the site were, both by design and by nature, more difficult than other areas of the site. This circumstance makes it difficult to compare the performance of different demonstrators that searched only portions of the site, if those portions of the site did not overlap. A second reason for measuring performance on the

entire site is that most of the demonstrators indicated in their proposals the ability and intention of searching the entire site in the allotted time. For these demonstrators, the fraction of the ordnance items emplaced on the entire site that were located in the allotted time is in some ways more meaningful for direct comparisons than the fraction of the ordnance items from a small part of the site. Demonstrators that indicated in their proposals that they would be unable to visit the entire site are so indicated.

1. Detection Capability

Detection capability is measured by a combination of four ratios. The overall detection ratio

$$R_{all} = \frac{|E|}{|B|} = \frac{\text{emplaced items detected}}{\text{total items emplaced}}$$

is a measure of the demonstrator's ability to find emplaced items, regardless of the demonstrator's ability to correctly identify the items.

The ordnance detection ratio

$$R_{ord} = \frac{|TP| + |MT|}{|BO|} = \frac{\text{total ordnance detected}}{\text{total ordnance emplaced}}$$

measures the demonstrator's ability to detect emplaced ordnance items, without regard to the demonstrator's ability to correctly identify the items as ordnance.

The nonordnance detection ratio

$$R_{non\text{-}ord} = \frac{|TN| + |FP|}{|BN|} = \frac{\text{total nonordnance detected}}{\text{total nonordnance emplaced}}$$

measures the demonstrator's ability to detect emplaced nonordnance items, without regard to the demonstrator's ability to correctly identify the items as nonordnance. In later sections, where alternative methods for matching baseline targets and demonstrator declarations are considered, the detection measures described here will be referred to as P_{match}.[1]

[1] The detection measures reported as R_x on $P_{match}(x)$, where x = all, ordnance or nonordnance, are equivalent to the detection probabilities reported in *Unexploded Ordnance Advanced Technology Demonstration Program at Jefferson Proving Ground (Phase 1)*, PRC, Inc.

The mistyped ordnance ratio

$$MR = \frac{|MT|}{|MT| + |TP|} = \frac{\text{ordnance declared as nonordnance}}{\text{ordnance items detected}}$$

measures a demonstrator's ability to distinguish ordnance from nonordnance, with emphasis on correctly designating ordnance. If all ordnance items detected are correctly identified as ordnance, $MR = 0$. Many demonstrators did not attempt to distinguish ordnance items from nonordnance items, placing all declarations in one category. For these demonstrators, *MR* was not a useful measure of performance.

For most demonstrators, the implications from the first three measures of detection capability were consistent.

The values reported by IDA for the above defined detection measures were calculated using the September 16, 1994 version of the target matching algorithm, which was tested by IDA. The emplaced target baseline set used in the calculations was from January 24, 1995. For the detection capabilities calculated on the area visited by the demonstrator, only the emplaced items and the demonstrator declarations within the grid cells that the demonstrator reported visiting were considered. There are small discrepancies between the numbers generated for this report and those reported in the *Unexploded Ordnance Advanced Technology Demonstration Program at Jefferson Proving Ground (Phase 1)* report prepared by PRC, Inc. for the Army Environmental Center (Ref. 2). The size of these discrepancies is consistent with the placement of a few demonstrator declarations into different areas of the Venn diagram. We believe this difference in placement is traceable to differences in one or more the above described procedures.

2. False Negative Rate

The false negative ratio

$$FNR = \frac{|FN|}{|FN| + |TP|} = \frac{\text{ordnance declarations not matched to baseline items}}{\text{ordnance declarations that are not matched to nonordnance}}$$

measures the demonstrator's ability to distinguish returns associated with natural or man-made clutter from returns associated with ordnance items in the ground. A low score here is good and high score is poor.

The area false alarm ratio

$$FAR = \frac{|FN| + |NF|}{Area} = \frac{\text{declarations not matched to baseline items}}{\text{area searched}}$$

measures the number of demonstrator declarations containing no known item that would have to be investigated per unit area in conducting site remediation. Since credit is given for an ordnance detection whether the declaration was typed as ordnance or not, both ordnance and nonordnance declarations contribute to this false alarm measure. The units of area are unspecified to prevent the demonstrators from back-calculating information about the number and types of ordnance items emplaced.

A false negative signal could arise from natural clutter, system noise, or items in the field that are unknown to the government. In an attempt to minimize the possibility of the latter, false alarms were investigated using the following procedure. Prior to the emplacement of any ordnance items, an Explosive Ordnance Disposal (EOD) team swept the test area and identified 72 anomalies. The locations of these anomalies were recorded. The declarations of the first six demonstrators to survey the site were compared with the anomalies found by the EOD team. Thirty of the anomalies were detected by two or more of these six demonstrators and, on that basis, were selected to be remediated after completion of all demonstrator surveys. Of the thirty anomalies remediated, two were mineral deposits, two were not investigated and therefore remain of unknown origin, and 26 were small nonordnance items near the surface. These thirty anomalies were added to the baseline nonordnance set (*BN*).

The addition of these anomalies to the nonordnance baseline set affects the measurements of overall probability of detection and probability of detecting nonordnance targets. The inclusion of these anomalies does not affect the calculation of probability of detection for ordnance items, which is the primary measure used in this evaluation of demonstrator performance. Demonstrator declarations matched to these anomalies are not counted as false alarms. Since the EOD team used magnetometers to identify the anomalies, their inclusion in the baseline may affect the results of the nonordnance detection and false negative rate for magnetometers differently than for other systems.

3. False Positive Rate

The false positive ratio

$$FPR = \frac{|FP|}{|FP| + |TN|} = \frac{\text{nonordnance detected and declared as ordnance}}{\text{nonordnance detected}}$$

measures the demonstrator's ability to distinguish ordnance from nonordnance, with emphasis on correctly designating nonordnance. A low score here is good and high score is poor.

4. Target Position and Accuracy

The ability of the demonstrator to accurately determine the position of a target is treated initially by selecting R_{crit}, the distance for determining a match between a demonstrator declaration and a baseline target using the target matching algorithm (i.e., if the position accuracy is worse than 2 m, a target is not counted as a match). For the targets that are determined to match baseline items, the average radial distance between the demonstrator declaration and baseline target ΔR_{xy} is calculated. For a statistical distribution of targets within the selected R_{crit}, this average radial distance should equal 2/3 R_{crit} or 1.3 m (4.3 feet). The average depth discrepancy ΔR_z is also calculated. ΔR_{xy} and ΔR_z are reported in feet.

5. Target Classification Capability

Seven classification ratios were created to measure the demonstrator's ability to correctly classify detected items. The form of the ratio is

$$R_i = \frac{|TP_{ii}|}{|E_{ix}|} = \frac{\text{correctly classified items of class } i}{\text{items of class } i \text{ detected by demonstrator}}$$

where i = single, multiple, bombs, projectiles, mortars, mines, and clusters. The first subscript refers to the baseline set, the second subscript refers to the demonstrator set, and x indicates all target classifications. For example, the classification capability for bombs would be measured by

$$R_{bombs} = \frac{|TP_{bombs,bombs}|}{|E_{bombs,x}|} = \frac{\text{bombs detected and correctly classified as bombs}}{\text{bombs detected}} \quad .$$

The classification ratios reported in this document differ from those found in the *Unexploded Ordnance Advanced Technology Demonstration Project at Jefferson Proving Ground (Phase I)* report prepared by PRC, Inc. (Ref. 2). Those in the PRC report are calculated as the ratio

$$R_{bombs}^{PRC} = \frac{|TP_{bombs,bombs}|}{|B_{bombs,x}|} = \frac{\text{bombs found and correctly classified as bombs}}{\text{bombs in the baseline}} \quad .$$

In this definition, the classification ratio depends not only on the demonstrator's ability to correctly classify targets, but also on the ability to detect them. We report separately the detection capabilities for each ordnance type emplaced for the demonstrators of ground based systems.

6. Survey Rate

The survey rate is measured as the portion of the site visited in the allotted 40 hours.

7. Cost

Costs, based on demonstrator firm fixed prices from the demonstrator proposals and area searched in the allotted time, are presented and discussed in Section 8.

Alternatives for Measuring Detection Capabilities

The measures of detection capability discussed above rely on one-to-one matches of demonstrator declarations with baseline items. Such a matching scheme assumes sensor resolution that is sufficient to separate closely spaced targets present in the baseline target set. If sensors do not have such resolution capabilities, the implications of using one-to-one matches for assessing demonstrator performance may be significant, as a large portion of the baseline targets are separated from their nearest neighbors by distances that are small compared to R_{crit}. If the resolution of the sensors is insufficient to separate two closely spaced targets, a more generous assessment of detection capability may be a measure of the number of baseline targets that would be detected if holes of radius R_{crit} were dug at all demonstrator declarations. Therefore, we define P_{near}

$$P_{near} = \frac{|N_n|}{|BO|} = \frac{\text{baseline ordnance items within } R_{crit} \text{ of a declaration}}{\text{total baseline ordnance items emplaced}} \quad ,$$

where N_n is the number of baseline items that have a demonstrator declaration within R_{crit} in horizontal distance.

The P_{near} measure will serve as an absolute upper limit for detection capability. If the demonstrator declared one detection at a location containing three baseline targets because the resolution of the sensor is not adequate to separate the targets, this measure will overstate detection capability. To illustrate, there are three ways of scoring one demonstrator declaration within R_{crit} of three closely spaced baseline targets, as shown in Fig. 2. The one-to-one matching method of the target matching algorithm, referred to as P_{match}, would conclude that the demonstrator located the one target that was the "best" match to the declaration, and that the other two targets were undetected. The P_{near} method counts one detection for each target within R_{crit} of the declaration and would conclude that the demonstrator located three of three possible baseline targets. The third possibility would

credit the demonstrator with detecting one *group* of targets, and not charge it for any missed detections. This measure is referred to as P_{group}.

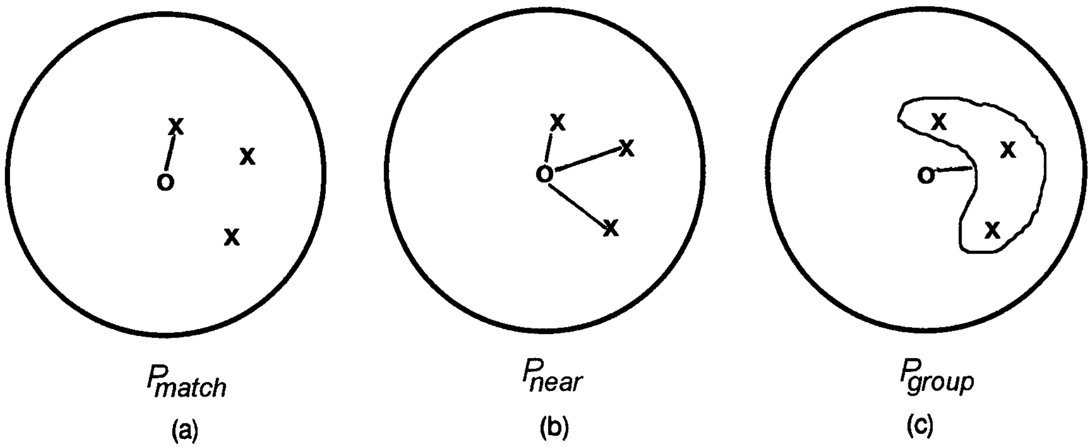


Figure 2. Three Methods of Scoring Three Baseline Within R_{crit} of One Demonstrator Declaration

The P_{match} scheme will likely understate the detection capability of the sensor, because it will determine that the sensor is unable to detect targets, when it was simply unable to resolve them from nearby targets. The P_{near} method will overstate detection capabilities, because it will give the sensor credit for detecting multiple items, where it only detected one group of items that it could not resolve into separate targets. The P_{group} method may give the most accurate representation of detection capability. However, it is difficult to rigorously compare different demonstrators, which will have different resolution capabilities, using this measure without knowing the details of the instrument resolution and the processing algorithms. Such information is required to make judgments about whether demonstrator declarations correspond to individual targets or to groups of targets. Further, for individual demonstrators it is likely that some declarations will fall into each of these two categories. However, this P_{group} measure can be approximated by assigning one detection for a group of targets captured by a single declaration, and defining the group as one baseline item. This approximation may also overstate demonstrator detection capabilities, as it relies on the assumption that all demonstrator declarations that can be matched to a group of targets indeed result from lack of resolution rather than lack of detection.

None of these three measures of detection capability directly corresponds to a probability of detection, P_d, in the usual sense. P_{match} and P_{group} can be used to put upper and lower bounds on P_d. P_{near}, which measures the fraction of baseline items that would

be recovered if a hole of radius R_{crit} was dug at each demonstrator declaration, is not a techically defendable measure, as it allows a single demonstrator declaration to be credited with multiple detections and, further, this ratio is strongly a function of the distribution of emplaced test articles. If the test articles are emplaced in groups with spacing that is close relative to R_{crit}, then many baseline items would be recovered serendipitously in holes of radius R_{crit} [see Fig. 2(c).]. P_{near} , more so than the other measures, is highly dependent on the configuration of the test at JPG in a way that unreasonably inflates detection probability. Thus, P_{near} cannot be used reliably to predict performance at other sites where the ordnance distribution is not known.

Using the Criteria for Demonstrator Evaluation

The seven evaluation criteria presented in the beginning of this section were selected prior to the demonstrations. At that time, it was believed in some quarters that detection capabilities would be high and false alarms would be sparse. Under such conditions, demonstrators with similar detection capabilities would have been separated by their ability to distinguish ordnance from nonordnance, location accuracy, ability to determine ordnance type, survey rate and cost. As a practical matter, the demonstrators did not exhibit capabilities that required this level of fidelity for performance evaluation, nor were the discrimination capabilities sufficient to make these tests meaningful. Therefore, demonstrator performance is evaluated primarily on the basis of probabilities of detection and false alarm rates.[2]

For all demonstrators on the 40-acre site, the maximum detection capabilities as estimated by the target matching algorithm have been lower than $P_{match} = 0.65$, and there were multiple false declarations per ordnance item detected for most demonstrators. For nearly half of the demonstrators, detection capabilities were below $P_{match} = 0.20$. Further, statistical analysis of the JPG data indicates that demonstrators with similar detection capabilities cannot be ranked ordinally with confidence. Therefore, demonstrators are binned on the basis of P_{match} and false alarm rate. Detection capability could have been binned on any of the three measures, P_{match}, P_{near}, or P_{group}, since the relative performance of the demonstrators does not change appreciably among the three measures. The bins for $P_{match}(ord)$ on area searched are (I) demonstrators that performed significantly better than the average demonstrator at JPG, (II) demonstrators that performed in the vicinity of the average demonstrator at JPG, and (III) demonstrators that performed

2 It should be noted that classification and accurate target position determination are government goals.

significantly worse than the average demonstrator at JPG. These categories were chosen based on an average P_{match} of 0.21 for all demonstrators on the 40-acre site.[3] The category of demonstrators with P_{match} in the vicinity of the average is between P_{match} = 0.16 and 0.26, i.e., 0.21 ± 0.05. The five percentage points capture the statistical uncertainties of most demonstrators.

$P_{match}(ord)$	Bin	Number of Demonstrators
> 0.26	I	8
0.16–0.26	II	3
< 0.16	III	9

The bins for false alarms, where P_{fa} is the fraction of the site covered by demonstrator false alarms, are as shown in the table below. P_{fa} ranged from a low of 0.004 for the demonstrator with the fewest false alarms to a high of 0.23 for the demonstrator with the most false alarms.

P_{fa}	Bin	Number of Demonstrators
0–0.01	I	5
0.01–0.03	II	9
0.03–0.08	III	4
> 0.08	IV	2

It is difficult to compare demonstrators with widely different P_{match} and false alarm rates. It is well recognized that these measures are inextricably linked, and standard approaches have been developed to quantify their relationship (Ref. 4). A variable *d* is introduced which measures the distance between the centroids of the target population and the background population measured in units of the standard deviation, assuming a gaussian noise distribution.[4] This *d* is a statistical parameter that indicates how well a particular measurement distinguishes signal from noise. Higher *d* values indicate better ability to separate signal from noise. It is possible to compute a *d* value from the demonstrator's P_{match} and P_{fa} values at a single point. This measure allows for comparison of demonstrators with widely different detection capabilities and false alarm

3 This average is the average over all demonstrators of the P_{match} computed on area searched, rather than the total number matches of all demonstrator detections divided by the total number of opportunities based on area searched.

4 For simplicity we assume the populations have the same standard deviation. If the standard deviations are different, an effective standard devition is given by the geometric of the two distributions.

rates. The statistical uncertainties discussed above preclude ranking the demonstrators one through twenty on the basis of this separation measure, so again a binning system, shown in the following table, is employed. The best *d* value for a system demonstrated on the 40-acre site at JPG is < 3. By way of comparison, a high performance surveillance radar would be designed to have a *d* of approximately 20 for the smallest targets of interest at the longest ranges of interest.

d[1]	Bin	Number of Demonstrators
> 2.0	I	2
1.5–2.0	II	5
1.0–1.5	III	5
< 1.0	IV	8

[1] A high *d* value is desirable.

There are some shortcomings to this approach. To accurately characterize the sensor, the entire receiver operating characteristic curve is desired. The mathematical model used here is imperfect even in the case of noise-limited radar detection, for which a gaussian distribution is a reasonable approximation of system noise. The clutter statistics, which currently dominate the false alarms for UXO detection, are not random and certainly not gaussian, and further are unlikely to be the same for the different sensors. Nevertheless, the approach provides a well defined measure, in use throughout the sensor community, that allows comparison among demonstrators. Furthermore, for the JPG demonstration, the *d* values are small, so deviation from gaussian statistics will be less important than for large values of *d*. Large values of *d* will be highly sensitive to the tails of the distributions, where differences from gaussian statistics will be greatest. Small values of *d*, indicating that the populations are not well separated, are determined from the more highly populated areas of the distributions, and will be more correctly estimated by a gaussian model.

Relative Difficulty on the 40-Acre Site

The 40-acre demonstration site was divided into four quadrants incorporating various concentrations and types of ordnance as well as nonordnance items that might generate false alarms. When demonstrator performance is assessed by quadrant, it is difficult to arrive at any compelling conclusions for two reasons. First, only three demonstrators searched the entire site. Second, dividing the data set into four pieces

increases the statistical uncertainties. Nevertheless, the following observations are supported by the demonstrator performance:

- The performance by quadrant is suggestive of one quadrant being easier than the others and one harder. The number of targets involved in this determination is sufficiently small that this apparent difference could be dominated by a few easy or a hard to find emplaced items or a modest difference in the natural clutter features of the quadrants.
- The same general pattern of relative difficulty emerges whether the determination is based on the declarations of only the three demonstrators that visited the entire site or the sum of all demonstrator declarations.
- Demonstrators that detected enough targets to allow a determination were in the same P_{match} bin for all four quadrants. In other words, demonstrators that did relatively poorly on the site as a whole did relatively poorly on all four quadrants, rather than well on the easy quadrant and terrible on the hard quadrants.
- The observed performance of individual demonstrators on multiple quadrants indicates that it is difficult to predict which conditions will challenge system capability with confidence. For example, in many cases a demonstrator that visited two or more quadrants performed better in the quadrant that was designed to be more difficult.

Evaluation Criteria and the Airborne Systems

None of the target declarations by any of the demonstrators of airborne systems can be attributed with confidence to a return from a single emplaced item. The number of matches scored by the airborne systems is consistent with the number of matches that would be expected from placing the same number of declarations at random. The false alarm measures reported are those used for the 40-acre site demonstrations. Given the insignificant number of matches, it is unclear what meaning to attribute to the false alarm measures.

The radial and depth accuracy are also computed and presented. The agreement of the radial accuracy with predictions assuming random placement of hits further supports the contention that no demonstrator declarations were actually caused by individual emplaced ordnance items. No significance is attributed to the computed depth accuracy. For the airborne demonstrators, the classification ratios and the probability of detection for the various ordnance types emplaced are disregarded. Since it appears that none of the

declarations can be associated with a return from a single target, these measures are not useful in characterizing the performance of the systems.

3. DESCRIPTION OF EVALUATION CRITERIA FOR DEMONSTRATORS OF REMEDIATION SYSTEMS

Demonstrators of remediation systems have been evaluated using the following criteria, listed in decreasing order of importance:

1. Ability to locate and excavate targets
2. Ability to assess the hazard posed by the target
3. Time for remediation
4. Volume of earth excavated
5. Cost

Remediation demonstrators were provided with the coordinates of the targets to be remediated. The excavators had to maneuver to the position and excavate the target(s). The navigation was sometimes assisted by flagging the targets. Excavation performance can be limited by the depth capabilities of the excavator.

The remediation excavators were equipped with remote cameras. Demonstrators were able to view some targets remotely. Other targets were not examined before they were moved by the excavators.

Time and volume of earth excavated were considered along with cost. The cost of the remediation will be directly affected by the time required to excavate each ordnance item, and also by the volume of earth that must be moved.

The data supplied by the demonstrators regarding each of these criteria are presented and discussed in the individual demonstrator reports. No attempt was made to quantify these measures for direct demonstrator comparison.

This page is intentionally left blank

4. SUMMARY OF DEMONSTRATORS' PERFORMANCE

Figure 3 summarizes the results of demonstrations on the 40-acre site with regard to probability of detection on the entire site, the probability of detection corrected for area searched, and the false alarm rate. Not all demonstrators were able to search the entire 40-acre site in the allotted time; therefore, Fig. 3*(b)* shows the probability of detection as measured only on the area visited by each demonstrator. In Fig. 3*(c)*, the false alarm rate is defined as the number of false alarms over the area searched. Most importantly, Fig. 3 highlights the importance of considering both probability of detection and false alarms in evaluating the performance of these systems.

The probability of detection is subject to various interpretations as discussed in Section 2, Description of Evaluation Criteria for Demonstrators of Detection Systems. Figure 4 compares the performance of demonstrators at the 40-acre site using three measures for ordnance detection: P_{match}, P_{near}, and P_{group}. As discussed, P_{near} is not a defendable measure of performance; thus, P_{group} and P_{match} can be viewed as upper and lower bounds on the performance of the detection systems at the 40-acre site. While these three methods result in different absolute measures of performance, with few exceptions, the relative performance of the demonstrators is retained regardless of the method used.

To distinguish demonstrators with high false alarm rates and high detection probability from those with low false alarm rates and low detection probability, we modeled the systems using *receiver operator curves* governed by gaussian noise. Using this model, the overall capability of the systems, including both detection and false alarms, can be compared by a single parameter, *d*. The higher *d*, the better the performance of the system. Figure 5 shows a plot of P_{match}(ordnance only) as measured only on the area searched against the false alarm rate. Overlaid on this plot are receiver operator curves with varying values of *d*. Demonstrators that intersect a curve with a high *d* value have better performance than those that intersect a curve with a low *d* value. Thus, these curves make it possible to compare demonstrators with data in substantially different areas of the figure.

As it was desirable to obtain an overall ranking of the demonstrators, Fig. 6 shows the probability of detection against the probability of false alarm with the performance "bins" identified. Three "bins" were established for probability of detection:

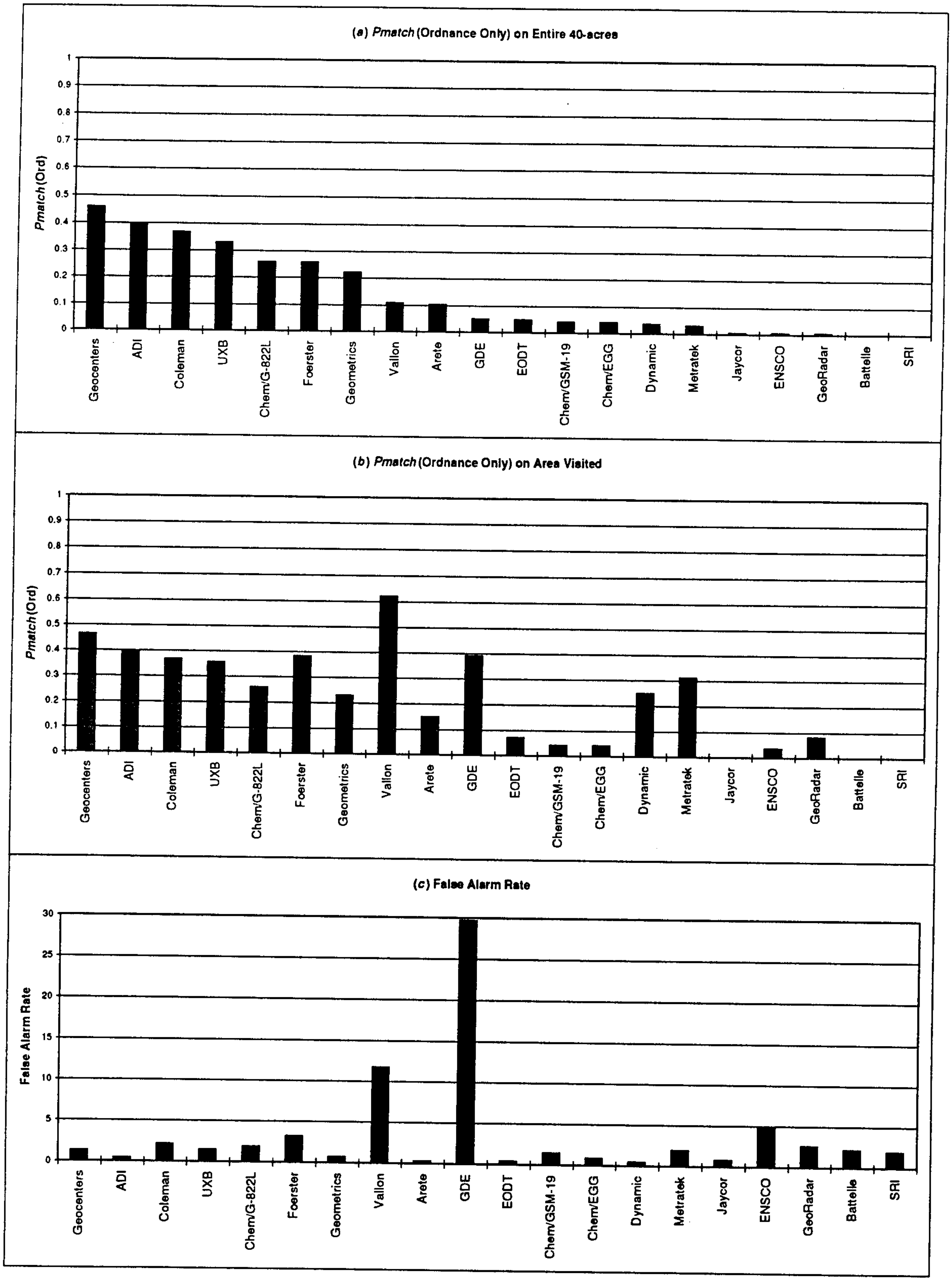


Figure 3. Demonstrator Performance on the 40-acre Site at Jefferson Proving Ground

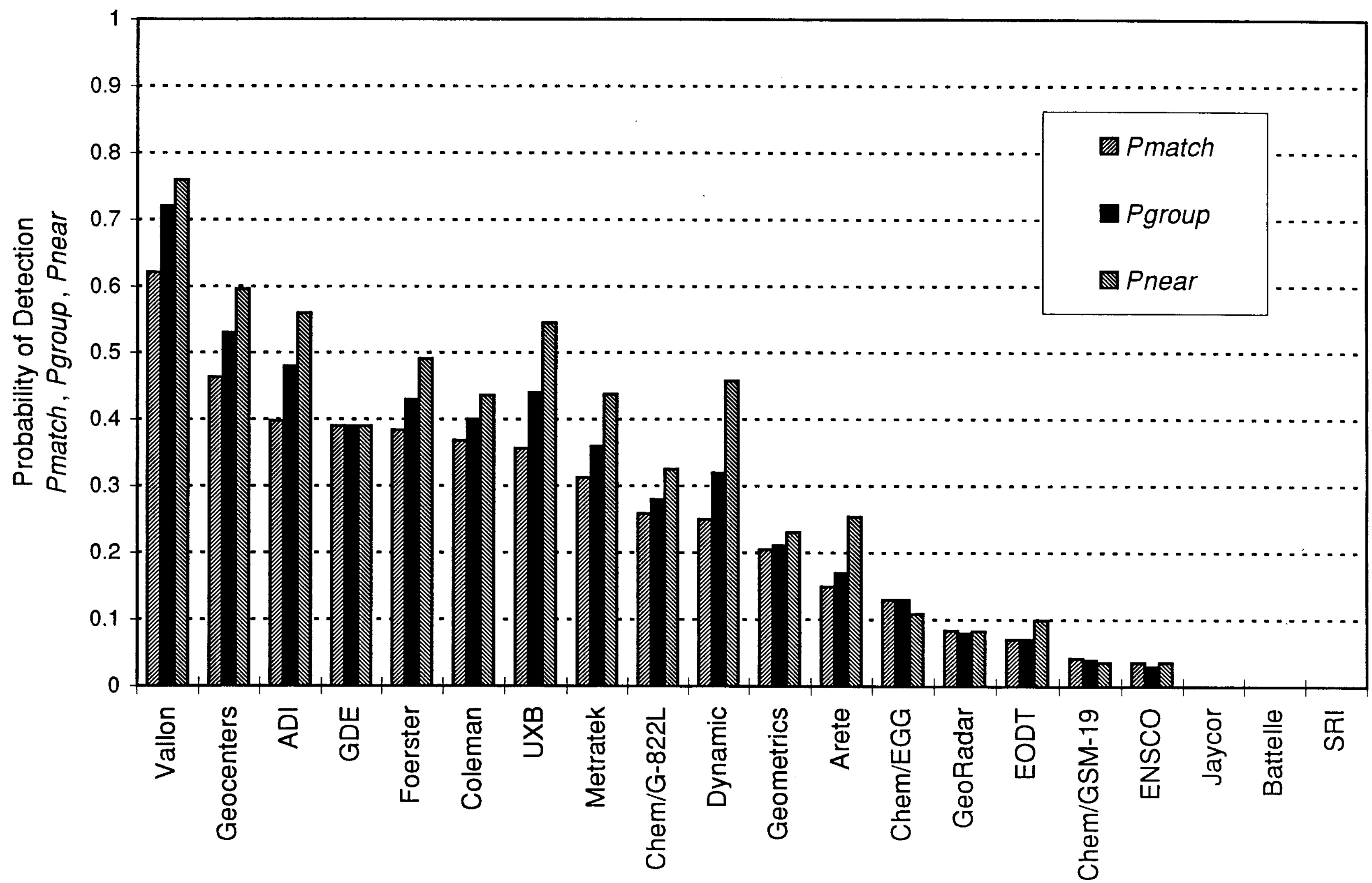
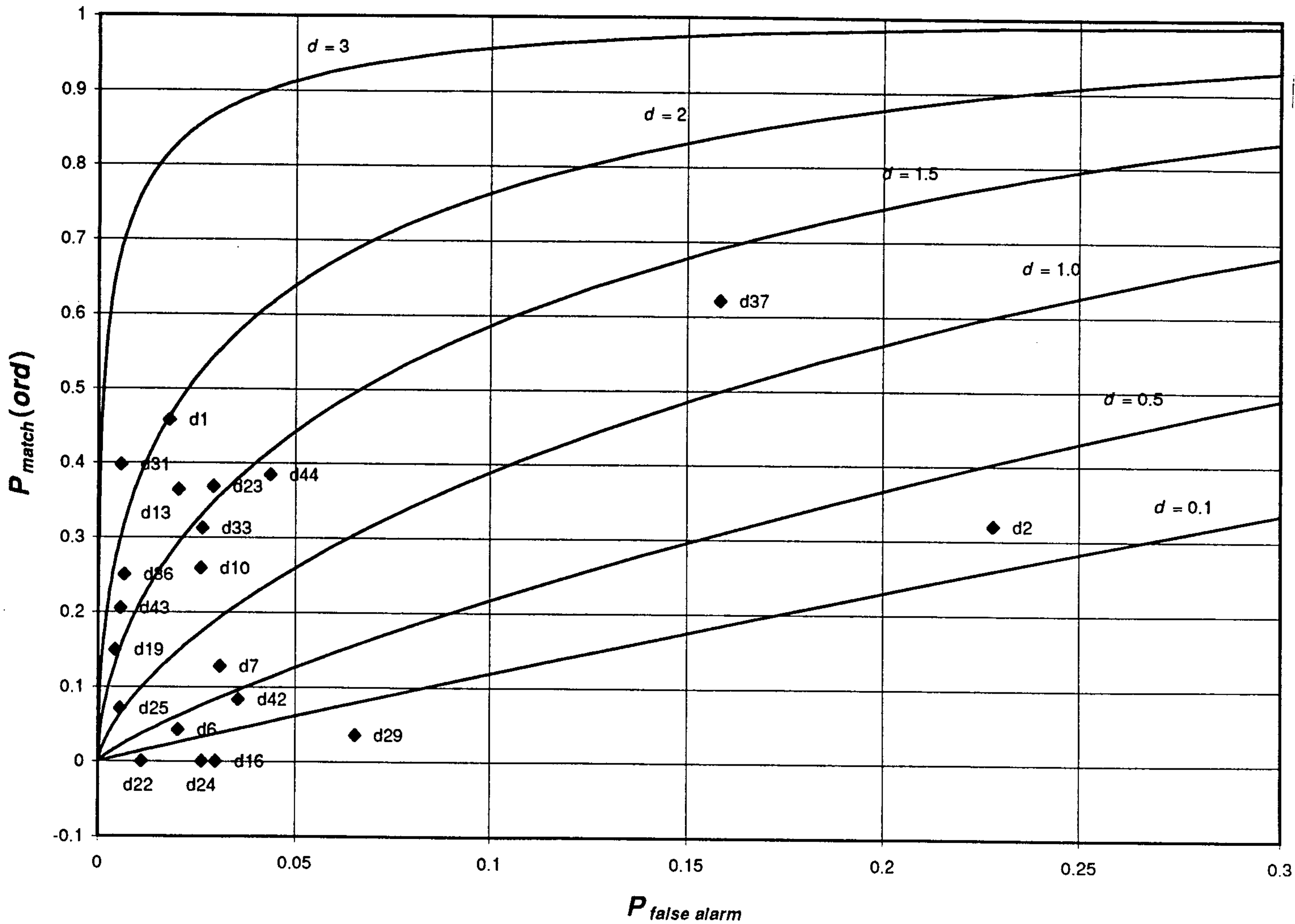


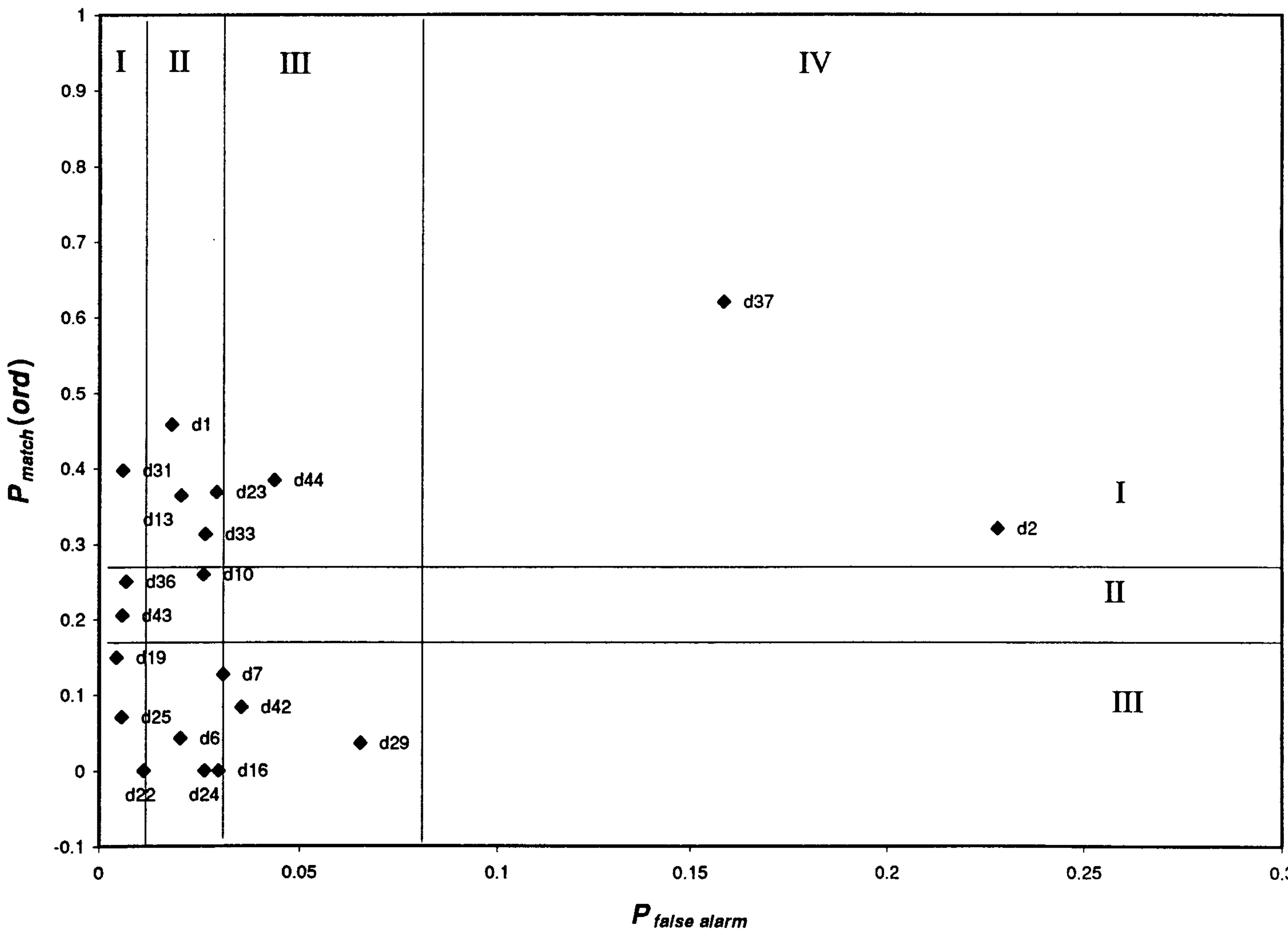
Figure 4. Comparison of Demonstrator Detection Capability as Measured by P_{match}, P_{group}, and P_{near}(ordnance only) on Area Searched



Key to Demonstrator Numbers

Number	Company
1	Geocenters
2	GDE
6	Chemrad(GSM-19)
7	Chemrad/EGG
10	Chemrad(G-822L)
13	UXB
16	Battelle/OSU
19	Arete
22	Jaycor
23	Coleman
24	SRI
25	EODT
29	ENSCO
31	ADI
33	Metratek
36	Dynamic
37	Vallon/SSPS
42	Geo-radar
43	Geometrics
44	Foerster

Figure 5. P_{match}**(ord) on Area Searched vs.** $P_{false\ alarm}$



Key to Demonstrator Numbers

Number	Company
1	Geocenters
2	GDE
6	Chemrad(GSM-19)
7	Chemrad/EGG
10	Chemrad(G-822L)
13	UXB
16	Battelle/OSU
19	Arete
22	Jaycor
23	Coleman
24	SRI
25	EODT
29	ENSCO
31	ADI
33	Metratek
36	Dynamic
37	Vallon/SSPS
42	Geo-radar
43	Geometrics
44	Foerster

Figure 6. P_{match}(ord) on Area Searched vs. $P_{false\ alarm}$ With Performance Bins Indicated

I. Significantly better than the average demonstrator at JPG.

II. In the vicinity of the average demonstrator.

III. Significantly worse than the average demonstrator.

The center "bin" was defined as the average performance capability, defined by P_{match}(ordnance only) on the area searched, plus or minus 5 percent. Thus, the average probability of detection was 21 percent and the center "bin" of performance includes all demonstrators with detection capabilities between 16 and 26 percent. "Bins" for performance as measured by the probability of false alarm were similarly defined. Table 1 is a summary of the demonstrators' performance on the 40-acre site.

Figure 7 shows the probability of detection against false alarm rate for demonstrations on the 80-acre site. For all demonstrated airborne systems, the probability of detection was not distinguishable from zero. False alarms are reported; however, no further measures of performance were pursued for the 80-acre site demonstrations.

Three remotely controlled remediation systems were demonstrated at JPG. These systems resulted in the successful recovery of 20 items and the partial recovery of 3 items. One of the systems successfully recovered all of 11 assigned targets; one system recovered 5 of 8 with partial success on the remaining 3; and the other system recovered 4 assigned targets and was unsuccessful in recovering the remaining 5 assigned.

Table 1. Summary of Demonstrator Performance on the 40-acre Site at Jefferson Proving Ground

Demonstrator	$P_{match}(ord)$ Bin	P_{fa} Bin	d Bin
ADI	I	I	I
Coleman	I	II	II
Geo-Centers	I	II	I
Metratek	I	II	III
UXB	I	II	II
Foerster	I	III	III
GDE	I	IV	IV
Security Search Products/Vallon	I	IV	III
Dynamic	II	I	II
Geometrics	II	I	II
Chemrad(G-822L)	II	II	III
Arete	III	I	II
EODT	III	I	III
Battelle/OSU	III	II	IV
Chemrad(GSM-19)	III	II	IV
Jaycor	III	II	IV
SRI	III	II	IV
Chemrad/EG&G	III	III	IV
ENSCO	III	III	IV
GeoRadar	III	III	IV

$P_{match}(ord)$ Bin	$P_{match}(ord)$ (on area searched)
I	> 0.26
II	0.16 – 0.26
III	< 0.16

P_{fa} Bin	$P_{false\ alarm}$ (defined as fraction of site area covered with false alarms, Rcrit = 2 m)
I	0 – 0.01
II	0.01 – 0.03
III	0.03 – 0.08
IV	> 0.08

d Bin	d (value of parameter used in receiver operator curve, gaussian assumption)
I	> 2
II	1.5 – 2
III	1 – 1.5
IV	< 1

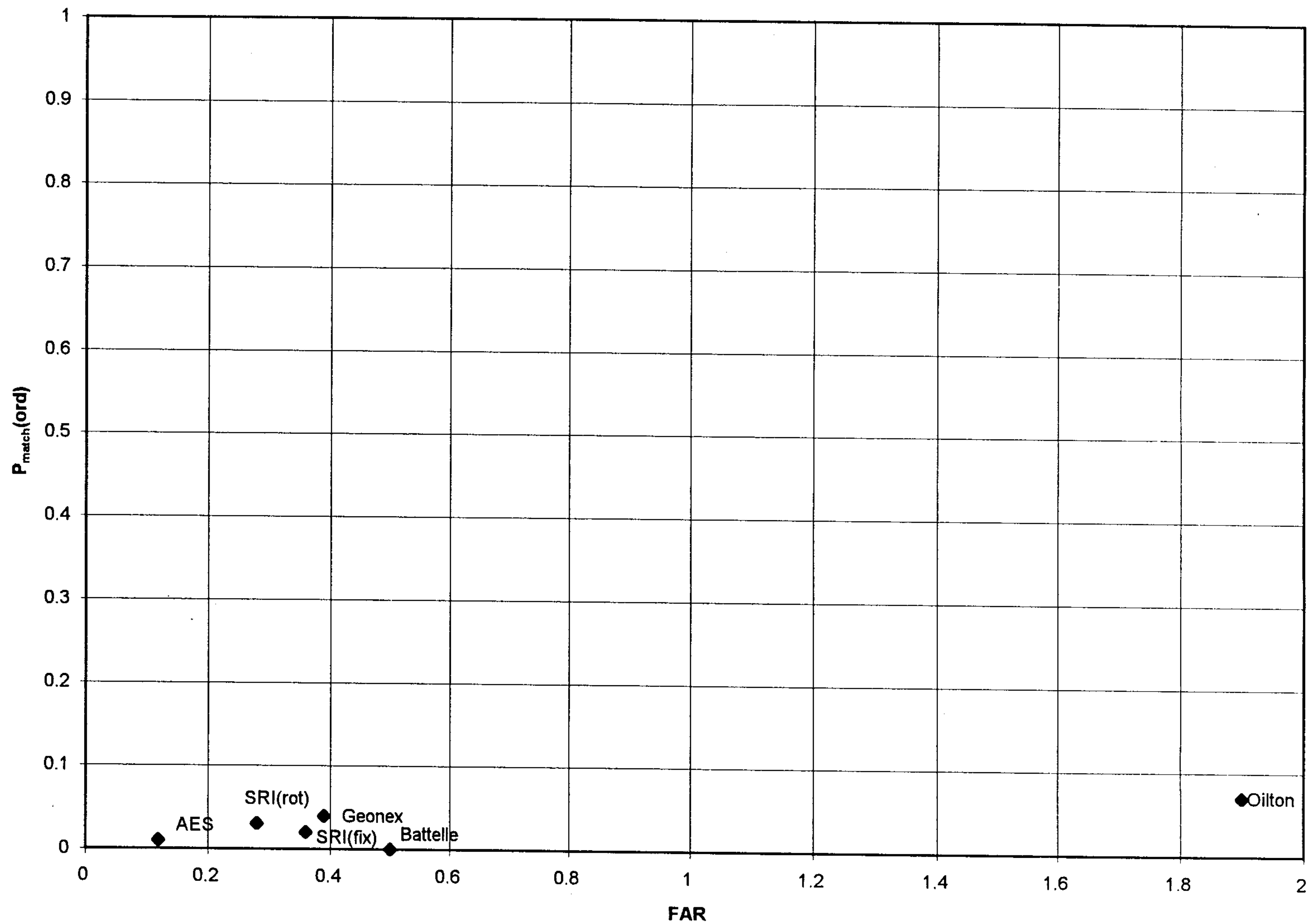


Figure 7. P_{match}**(ord) on Area Searched vs. False Alarm Rate for Airborne Systems**

5. 40-ACRE SITE EVALUATIONS

This page is intentionally left blank

EVALUATION OF ARETÉ ENGINEERING TECHNOLOGIES CORP. 40-ACRE SITE

I. DEMONSTRATOR RESULTS

	Entire Site	Area Searched[1]
Detection Capability		
overall detection ratio, R_{all}	0.11	0.17
ordnance detection ratio, R_{ord}	0.10	0.15
nonordnance detection ratio, R_{nonord}	0.13	0.21
mistyped ordnance ratio, *MR*	0.06	0.06
False Positive Ratio, *FPR*	0.90	0.90
False Negative Ratio, *FNR*	0.69	0.69
False Alarm Rate, *FAR*	0.21	0.32
Radial Accuracy, ΔR_{xy} *(feet)*	1.74	1.74
Depth Accuracy, ΔR_z *(feet)*	1.24	1.24
Target Classification Capabilities		
single targets	0.882	0.882
multiple targets	*	*
bombs	0.667	0.667
projectiles	0.833	0.833
mortars	0	0
mines	*	*
clusters	*	*

[1] The portion of the site searched by this demonstrator may not be representative of the overall difficulty of the entire 40-acre site.

* Ratio is undefined (i.e., zero denominator). A classification ratio is undefined if no targets of that class are located by the demonstrator.

Survey Rate

The demonstrator searched approximately 25 acres of the 40-acre site within the allotted 1 week. The proposal gave the survey rate as 10 acres per day for the hand-held

system. The demonstrator report comments on the difficult operating conditions at JPG, but does not indicate that this is the reason the survey rate fell short of the demonstrator's expectations.

II. DISCUSSION OF DEMONSTRATOR PERFORMANCE

Areté Engineering Technologies used Geonics induction coil and Schonstedt gradiometer sensors connected to a man-portable GeoDAPS control system for detection, and a Trimble differential GPS system for navigation. Both detectors have ranges up to 25 feet or more, depending on the size of the buried object. The demonstrator report comments on the operating difficulties caused by muddy and slippery conditions, combined with uneven ground, ditches, and sinkholes. The report states that these conditions improved somewhat during the course of the week.

The demonstrator surveyed approximately 25 acres of the 40-acre site. When the demonstrator's detection capability and false alarm rates are recalculated to include only the portion of the site searched, the results are as follows:

$$P_{match} = 0.17$$

$$P_{match}\,(\text{ord}) = 0.15$$

$$FAR = 0.32$$

Corrections for random hits by this demonstrator reduced the uncorrected P_{match} = 0.15 to a corrected P_{match} =0.14, where P_{match} refers to the ordnance detection capability on only the area searched.

Detection capabilities were calculated for the various ordnance types emplaced. Only the targets in the area searched were considered, and credit was given for a detection regardless of classification ability, when calculating the detection capabilities reported in this paragraph. This demonstrator's best performance was in the detection of projectiles, with P_{match} = 0.32. This demonstrator's worst performance was in detecting mines and cluster targets, with zero detections of either target type. Since the sensors used by this demonstrator can detect only metal objects, the inability to detect plastic mines is to be expected. The performance of this demonstrator was best at intermediate depths. When the targets are grouped as those buried less than 3 feet, between 3 and 6 feet, and more than 6 feet deep, P_{match} was 0.14, 0.26, and 0.12, respectively. The demonstrator was best able to detect medium size targets, with P_{match} = 0.41, and detected small and large targets at P_{match} = 0.02 and P_{match} = 0.25, respectively.

The detection capabilities calculated above are all based on P_{match}, the probability of a one-to-one match between a demonstrator declaration and an emplaced target. This method of determining detection probability assumes a high degree of sensor resolution, which may not be justified if a sensor cannot resolve signals resulting from two (or more) closely spaced objects. Two alternative measures of detection capability are also considered. For demonstrator declarations with multiple baseline items within R_{crit}, P_{group} credits the demonstrator with detecting a single group of targets, and does not penalize for any missed detections. P_{near} measures the probability that an item will be found if holes of radius R_{crit} are dug on all demonstrator declarations, allowing multiple baseline targets to be matched to a single demonstrator declaration. The following table compares detection capability calculated using these three measures, where P_{group} and P_{match} represent upper and lower bounds on the probability of detection, P_d, achieved on this demonstration. The false alarm rate, *FAR*, will be the same for all three measures; the false negative ratio will be different. Also presented in the table are P_{match} and P_{near} calculated with the plastic mines removed from the baseline. All detection measures are for ordnance items only.

Calculation method	Value
P_{match}[1]	0.15
P_{match}*(no mines)*	0.18
P_{group}	0.17
P_{near}	0.25
P_{near} *(no mines)*	0.32

[1] For this demonstration, the correction to detection capabilities for random matches was small (< 0.03).

The following tables provide information about specific demonstrator declarations and baseline items which may be of use to the demonstrator in assessing target identification capabilities for future operations, or for making improvements in the sensor or the data processing algorithms. The correct locations and types are provided for 17 baseline items that were dug up during or shortly after the remediation demonstrations. Five of these baseline items are isolated, individual objects. The remaining 12 are six closely spaced pairs of items. These items have either been removed from the site or were emplaced again at new locations. Three demonstrator declarations that were not within 2 meters of a baseline item are also provided.

Locations and Types of Selected Baseline Targets				
Type	Size	X[a]	Y	Depth
Bomb	Large	2104200.6	14139531.2	8.0
Bomb	Large	2105272.6	14139550.2	5.6
Bomb	Large	2105272.5	14139555.1	2.9
Projectile	Medium	2105087.4	14139006.9	3.4
Projectile	Medium	2105105.8	14139336.8	2.7
Projectile	Small	2105134.5	14139561.0	6.1
Projectile	Medium	2105326.1	14139555.1	2.0
Projectile	Medium	2105014.3	14139251.5	3.3
Mortar	Small	2105139.5	14139036.9	3.1
Mortar	Small	2105327.9	14139554.6	1.2
Mortar	Small	2104978.4	14139240.2	1.3
Mortar	Small	2104976.0	14139229.1	0.8
Mortar	Small	2104976.7	14139229.3	0.9
Mortar	Small	2104013.6	14138535.2	1.1
Mortar	Small	2104017.5	14138533.9	0.9
NonOrd	Large	2105014.7	14139251.1	0.9
NonOrd	Large	2105139.8	14139037.1	1.1
[a] All dimensions are in feet.				

Demonstrator Declarations Not Within 2 m of a Baseline Item	
2001[a]	No known baseline item within 2 m of this location.
2006	No known baseline item within 2 m of this location.
3024	No known baseline item within 2 m of this location.
[a] Target numbers are from the demonstrator numbering scheme.	

III. SUMMARY

The purpose of the 1994 Jefferson Proving Ground demonstration was to evaluate the state of the art in UXO detection and remediation technology. For all detection demonstrators, the maximum detection capabilities as measured by the target matching algorithm have been lower than $P_{match} = 0.65$ and there have been multiple false declarations per ordnance item detected. Further, statistical analysis of the JPG data

indicates that demonstrators with similar detection capabilities cannot be ordinally ranked with confidence. Therefore, demonstrators are binned on the basis of P_{match}, false alarm rate, and a separation criteria d, which combines measures of P_{match} and P_{fa} into a single variable and allows for comparison of demonstrators with widely different detection capabilities and false alarm rates. In this assessment, Aretè performed significantly worse than the average JPG demonstrator in terms of P_{match}, in the best bin of four bins in false alarms, and in the second best of four bins in the separation measure.

This page is intentionally left blank

EVALUATION OF AUSTRALIAN DEFENCE INDUSTRIES 40-ACRE SITE

I. DEMONSTRATOR RESULTS

	Entire Site	Area Searched[1]
Detection Capability		
overall detection ratio, R_{all}	0.48	0.48
ordnance detection ratio, R_{ord}	0.45	0.45
nonordnance detection ratio, R_{nonord}	0.54	0.54
mistyped ordnance ratio, MR	0.00	0.00
False Positive Ratio, FPR	0.93	0.93
False Negative Ratio, FNR	0.75	0.75
False Alarm Rate, FAR	2.8	2.8
Radial Accuracy, ΔR_{xy} *(feet)*	2.2	2.2
Depth Accuracy, ΔR_z *(feet)*	1.5	1.5
Target Classification Capabilities		
single targets	1.0	1.0
multiple targets	0.0	0.0
bombs	1.0	1.0
projectiles	0.85	0.85
mortars	0.0	0.0
mines	*	*
clusters	0.0	0.0

[1] These columns are identical because this demonstrator searched the entire 40-acre site.

* Ratio is undefined (i.e., zero denominator). The classification ratios undefined because no targets of the class were located.

Survey Rate

The demonstrator searched the entire 40-acre site within the allotted 1 week. The proposal gave survey rates as 10 acres per day for hand-held magnetometer and 20–40 acres per day for the surface towed magnetometer.

II. DISCUSSION OF DEMONSTRATOR PERFORMANCE

ADI used hand-held and surface towed GT-TM4 magnetometers to survey the site and a GT odometer, rope, and tape for navigation. The sensor used is not able to determine the orientation of a projectile in the ground. The computer aided interpretation of the magnetometer data provides a list of position, depth, and approximate mass of targets located. ADI proposed to demonstrate two independent magnetic surveys, one hand-held and one surface-towed to demonstrate comparative speed and data quality. The demonstrator proposed to demonstrate a GPR system as well, but did not do so because the soil conductivity was too high for reliable GPR performance. ADI determined at the time they were on the site that the conductivity was 10 times higher than that reported by the government/PRC.

Regarding the false alarm ratios, the demonstrator report states that any metal object greater than or equal to 100g mass is reported due to safety considerations. The demonstrator is willing to accept high *FNR* and *FPR* because it believes that all ferrous objects should be remediated, since it is impossible to say with 100% assurance that an object is not live ordnance based on its magnetic signature.

A number of the items that the demonstrator classified as ordnance items were indicated in the comment field as an old fence line or geologic anomaly. The P_{match}, *FNR*, *FPR*, and *FAR* for this demonstrator were recalculated after taking out all such declarations. The adjusted values are:

$$P_{match} = 0.41$$

$$P_{match}(\text{ord}) = 0.40$$

$$FNR = 0.53$$

$$FPR = 0.97$$

$$FAR = 0.43$$

where P_{match} is the overall detection ratio. The demonstrator's *FNR* and *FAR* are significantly reduced. P_{match} is also slightly reduced, as some of the declarations indicated as fence or anomaly were, in fact, matched to emplaced items. The measures used for the remainder of this evaluation are calculated with these declarations removed from the demonstrator data set.

Corrections for random hits by this demonstrator, with the fence and anomalies removed, reduced the uncorrected $P_{match} = 0.40$ to $P_{match} = 0.39$, for ordnance items.

Detection capabilities were calculated for the various ordnance types emplaced. Credit was given for a detection regardless of classification ability, when calculating the detection capabilities reported in this paragraph. This demonstrator's best performance was in detection of bombs, with $P_{match} = 0.84$. Detection ability for projectiles and mortars were $P_{match} = 0.51$ and 0.32, respectively. This demonstrator's worst performance was in detecting mines, with zero detections. Since the sensor used by this demonstrator can detect only metal objects, the inability to detect plastic mines is to be expected. The performance of this demonstrator was best at the lowest depths. When the targets are grouped as those buried less than 3 feet, between 3 and 6 feet, and more than 6 feet deep, P_{match} was 0.38, 0.37, and 0.68, respectively. The demonstrator was best able to detect large targets, with $P_{match} = 0.58$, and detected small and medium targets at $P_{match} = 0.27$ and 0.49, respectively.

The detection capabilities calculated above are all based on P_{match}, the probability of a one-to-one match between a demonstrator declaration and an emplaced target. This method of determining detection probability assumes a high degree of sensor resolution, which may not be justified if a sensor cannot resolve signals resulting from two (or more) closely spaced objects. Two alternative measures of detection capability are also considered. For demonstrator declarations with multiple baseline items within R_{crit}, P_{group} credits the demonstrator with detecting a single group of targets, and does not penalize for any missed detections. P_{near} measures the probability that an item will be found if holes of radius R_{crit} are dug on all demonstrator declarations, allowing multiple baseline targets to be matched to a single demonstrator declaration. The following table compares detection capability calculated using these three measures, where P_{group} and P_{match} represent upper and lower bounds on the probability of detection, P_d, achieved on this demonstration. The false alarm rate, *FAR*, will be the same for all three measures; the false negative ratio will be different. Also presented in the table are P_{match} and P_{near} calculated with the plastic mines removed from the baseline. All detection measures are for ordnance items only.

Calculation method	Value
P_{match}[1]	0.40
P_{match}(no mines)	0.46
P_{group}	0.48
P_{near}	0.56
P_{near} (no mines)	0.65

[1] For this demonstration, the correction to detection capabilities for random matches was small (< 0.03).

The following tables provide information about specific demonstrator declarations and baseline items which may be of use to the demonstrator in assessing target identification capabilities for future operations, or for making improvements in the sensor or the data processing algorithms. The correct locations and types are provided for 17 baseline items that were dug up during or shortly after the remediation demonstrations. Five of these baseline items are isolated, individual objects. The remaining 12 are six closely spaced pairs of items. These items have either been removed from the site or were emplaced again at new locations. Three demonstrator declarations that were not within 2 meters of a baseline item are also provided.

Locations and Types of Selected Baseline Targets				
Type	Size	X[a]	Y	Depth
Bomb	Large	2104200.6	14139531.2	8.0
Bomb	Large	2105272.6	14139550.2	5.6
Bomb	Large	2105272.5	14139555.1	2.9
Projectile	Medium	2105087.4	14139006.9	3.4
Projectile	Medium	2105105.8	14139336.8	2.7
Projectile	Small	2105134.5	14139561.0	6.1
Projectile	Medium	2105326.1	14139555.1	2.0
Projectile	Medium	2105014.3	14139251.5	3.3
Mortar	Small	2105139.5	14139036.9	3.1
Mortar	Small	2105327.9	14139554.6	1.2
Mortar	Small	2104978.4	14139240.2	1.3
Mortar	Small	2104976.0	14139229.1	0.8
Mortar	Small	2104976.7	14139229.3	0.9
Mortar	Small	2104013.6	14138535.2	1.1
Mortar	Small	2104017.5	14138533.9	0.9
NonOrd	Large	2105014.7	14139251.1	0.9
NonOrd	Large	2105139.8	14139037.1	1.1

[a] All dimensions are in feet.

Demonstrator Declarations Not Within 2 m of a Baseline Item	
286[a]	No known baseline item within 2 m of this location.
116	No known baseline item within 2 m of this location.
598	No known baseline item within 2 m of this location.
[a] Target numbers are from the demonstrator numbering scheme.	

III. SUMMARY

The purpose of the 1994 Jefferson Proving Ground demonstration was to evaluate the state of the art in UXO detection and remediation technology. For all detection demonstrators, the maximum detection capabilities as measured by the target matching algorithm have been lower than P_{match} = 0.65 and there have been multiple false declarations per ordnance item detected. Further, statistical analysis of the JPG data indicates that demonstrators with similar detection capabilities cannot be ordinally ranked with confidence. Therefore, demonstrators are binned on the basis of P_{match}, false alarm rate, and a separation criteria *d*, which combines measures of P_{match} and P_{fa} into a single variable and allows for comparison of demonstrators with widely different detection capabilities and false alarm rates. In this assessment, ADI performed significantly better than the average JPG demonstrator in terms of P_{match}, in the top bin of four bins in false alarms, and in the top bin of four bins in the separation measure, when declarations indicated as fence or geologic anomaly are removed from the demonstrator data set.

This page is intentionally left blank

EVALUATION OF BATTELLE/OSU
40-ACRE SITE

I. DEMONSTRATOR RESULTS

	Entire Site	Area Searched[1]
Detection Capability		
overall detection ratio, R_{all}	0.004	0.06
ordnance detection ratio, R_{ord}	0.0	0.0
nonordnance detection ratio, R_{nonord}	0.013	0.17
mistyped ordnance ratio, *MR*	*	*
False Positive Ratio, *FPR*	1.0‡	1.0‡
False Negative Ratio, *FNR*	1.0	1.0
False Alarm Rate, *FAR*	0.13	2.2
Radial Accuracy, ΔR_{xy} *(feet)*	1.6	1.6
Depth Accuracy, ΔR_z *(feet)*	0.5	0.5
Target Classification Capabilities		
single targets	*	*
multiple targets	*	*
bombs	*	*
projectiles	*	*
mortars	*	*
mines	*	*
clusters	*	*

[1] The portion of the site searched by this demonstrator may not be representative of the overall difficulty of the entire 40-acre site.

* Ratio is undefined (i.e., zero denominator). A classification ratio is undefined if no targets of that class are located by the demonstrator.

‡ The values shown here are a consequence of the demonstrator declaring all items as ordnance.

Survey Rate

The demonstrator searched approximately 2.3 acres of the 40-acre site within the allotted 1 week. Equipment problems caused the loss of 1 day's data. The demonstrator states in the proposal the intention to use the allotted 5 days for the demonstration. No indication of survey rate is given, except that the survey rate is slower than will ultimately be available, because the demonstrator will be using an experimental, laboratory version of the radar system.

II. DISCUSSION OF DEMONSTRATOR PERFORMANCE

Battelle and the Ohio State University used a surface towed ground penetrating radar for detection, and rope/tape/odometer system for navigation. The demonstrator report comments that the soil at JPG is high in clay content, and is, therefore, as noted in the proposal, a very challenging environment for radar systems. The demonstrator further comments that while the detection capabilities at JPG were poor, the demonstration provided information that will be useful in designing an antenna configuration that is more suited to UXO detection.

This demonstrator did not attempt to determine whether items were ordnance or nonordnance, typing all detections as ordnance items. Because no attempt was made to distinguish ordnance from nonordnance, several of the ratios used in this evaluation are not useful for characterizing this demonstrator. For example, $MR = 0$ not because the demonstrator made no errors in separating ordnance from nonordnance; rather, because no items were typed as nonordnance, no ordnance items could be misidentified as nonordnance. $FPR = 1.0$ since any nonordnance items found were typed as ordnance. Of particular note for this demonstrator is the false negative ratio of 1.0. As a practical matter, if all the ordnance declarations of this demonstrator were investigated, some holes would contain emplaced nonordnance items and the remaining holes dug for remediation would not contain any of the ordnance items known by the government to be present.

The demonstrator surveyed approximately 2.3 acres of the 40-acre site. When the demonstrator's detection capability and false alarm rates are recalculated to include only the portion of the site searched, the results are as follows:

$$P_{match} = 0.06$$

$$P_{match}(\text{ord}) = 0.0$$

$$FAR = 2.2$$

Corrections for random hits by this demonstrator reduced the uncorrected P_{match} = 0.0 to a corrected P_{match} = – 0.03, where P_{match} refers to the ordnance detection capability on only the area searched.

This demonstrator found no ordnance items, making the classification ratios and measures of detection capability in terms of target classification, size, and depth meaningless.

The detection capabilities calculated above are all based on P_{match}, the probability of a one-to-one match between a demonstrator declaration and an emplaced target. This method of determining detection probability assumes a high degree of sensor resolution, which may not be justified if a sensor cannot resolve signals resulting from two (or more) closely spaced objects. Two alternative measures of detection capability are also considered. For demonstrator declarations with multiple baseline items within R_{crit}, P_{group} credits the demonstrator with detecting a single group of targets, and does not penalize for any missed detections. P_{near} measures the probability that an item will be found if holes of radius R_{crit} are dug on all demonstrator declarations, allowing multiple baseline targets to be matched to a single demonstrator declaration. The following table compares detection capability calculated using these three measures, where P_{group} and P_{match} represent upper and lower bounds on the probability of detection, P_d, achieved on this demonstration. The false alarm rate, *FAR*, will be the same for all three measures; the false negative ratio will be different. Also presented in the table are P_{match} and P_{near} calculated with the plastic mines removed from the baseline. All detection measures are for ordnance items only.

Calculation method	Value
P_{match}	0.0
P_{match}(no mines)	0.0
P_{group}	0.0
P_{near}	0.0
P_{near} (no mines)	0.0

The following tables provide information about specific demonstrator declarations and baseline items which may be of use to the demonstrator in assessing target identification capabilities for future operations, or for making improvements in the sensor or the data processing algorithms. The correct locations and types are provided for 17 baseline items that were dug up during or shortly after the remediation demonstrations. Five of these baseline items are isolated, individual objects. The remaining 12 are six closely spaced pairs of items. These items have either been removed from the site or were

emplaced again at new locations. Three demonstrator declarations that were not within 2 meters of a baseline item are also provided.

Locations and Types of Selected Baseline Targets				
Type	Size	X[a]	Y	Depth
Bomb	Large	2104200.6	14139531.2	8.0
Bomb	Large	2105272.6	14139550.2	5.6
Bomb	Large	2105272.5	14139555.1	2.9
Projectile	Medium	2105087.4	14139006.9	3.4
Projectile	Medium	2105105.8	14139336.8	2.7
Projectile	Small	2105134.5	14139561.0	6.1
Projectile	Medium	2105326.1	14139555.1	2.0
Projectile	Medium	2105014.3	14139251.5	3.3
Mortar	Small	2105139.5	14139036.9	3.1
Mortar	Small	2105327.9	14139554.6	1.2
Mortar	Small	2104978.4	14139240.2	1.3
Mortar	Small	2104976.0	14139229.1	0.8
Mortar	Small	2104976.7	14139229.3	0.9
Mortar	Small	2104013.6	14138535.2	1.1
Mortar	Small	2104017.5	14138533.9	0.9
NonOrd	Large	2105014.7	14139251.1	0.9
NonOrd	Large	2105139.8	14139037.1	1.1
[a] All dimensions are in feet.				

Demonstrator Declarations Not Within 2 m of a Baseline Item	
1[a]	No known baseline item within 2 m of this location.
5	No known baseline item within 2 m of this location.
210	No known baseline item within 2 m of this location.
[a] Target numbers are from the demonstrator numbering scheme.	

III. SUMMARY

The purpose of the 1994 Jefferson Proving Ground demonstration was to evaluate the state of the art in UXO detection and remediation technology. For all detection demonstrators, the maximum detection capabilities as measured by the target matching

algorithm have been lower than $P_{match} = 0.65$ and there have been multiple false declarations per ordnance item detected. Further, statistical analysis of the JPG data indicates that demonstrators with similar detection capabilities cannot be ordinally ranked with confidence. Therefore, demonstrators are binned on the basis of P_{match}, false alarm rate, and a separation criteria d, which combines measures of P_{match} and P_{fa} into a single variable and allows for comparison of demonstrators with widely different detection capabilities and false alarm rates. In this assessment, Battelle performed significantly worse than the average JPG demonstrator in terms of P_{match}, in the second of four bins in false alarms, and in the worst of four bins in the separation measure.

This page is intentionally left blank

EVALUATION OF CHEMRAD/EG&G
40-ACRE SITE

I. DEMONSTRATOR RESULTS

	Entire Site	Area Searched[1]
Detection Capability		
overall detection ratio, R_{all}	0.05	0.13
ordnance detection ratio, R_{ord}	0.04	0.13
nonordnance detection ratio, R_{nonord}	0.05	0.14
mistyped ordnance ratio, MR	0.0‡	0.0‡
False Positive Ratio, FPR	1.0‡	1.0‡
False Negative Ratio, FNR	0.96	0.96
False Alarm Rate, FAR	0.91	2.28
Radial Accuracy, ΔR_{xy} *(feet)*	5.38	5.38
Depth Accuracy, ΔR_z *(feet)*	3.47	3.47
Target Classification Capabilities		
single targets	1.0	1.0
multiple targets	0.0	0.0
bombs	0.5	0.5
projectiles	1.0	1.0
mortars	0.0	0.0
mines	*	*
clusters	0.0	0.0

[1] The portion of the site searched by this demonstrator may not be representative of the overall difficulty of the entire 40-acre site.

* Ratio is undefined (i.e., zero denominator). A classification ratio is undefined if no targets of that class are detected by the demonstrator.

‡ The values shown here are a consequence of the demonstrator declaring all items as ordnance.

Survey Rate

The demonstrator searched approximately 16 acres of the 40-acre site within the allotted 1 week. The proposal gave the survey rate as 40 acres per 4 days.

II. DISCUSSION OF DEMONSTRATOR PERFORMANCE

Chemrad/EG&G used a Gulf Applied ground penetrating radar and a Pulse Tech induction coil sensor to survey the site. An acoustic USRADS surveying tool was used for navigation. The accuracy of the navigation tool is 6 inches.

This demonstrator did not attempt to determine whether items were ordnance or nonordnance, typing all detections as ordnance items. Because no attempt was made to distinguish ordnance from nonordnance, several of the ratios used in this evaluation are not useful for characterizing this demonstrator. For example, $MR = 0$ not because the demonstrator made no errors in separating ordnance from nonordnance; rather, because no items were typed as nonordnance, no ordnance items could be misidentified as nonordnance. $FPR = 1.0$ since any nonordnance items found were typed as ordnance. Of particular note for this demonstrator is the false negative ratio of 0.96. As a practical matter, if all the ordnance declarations of this demonstrator were investigated, some holes would contain emplaced nonordnance items, and 96% of the remaining holes dug would not contain any of the ordnance items known by the government to be present.

The demonstrator surveyed approximately 16 acres of the 40-acre site. When the demonstrator's detection capability and false alarm rates are recalculated to include only the portion of the site searched, the results are as follows:

$$P_{match} = 0.13$$

$$FAR = 2.28$$

Corrections for random hits by this demonstrator reduced the uncorrected $P_{match} = 0.13$ to a corrected $P_{match} = 0.10$, where P_{match} refers to the ordnance detection capability on only the area searched.

Detection capabilities were calculated for the various ordnance types emplaced. Only the items in the area searched are considered for this analysis, and credit is given for a detection regardless of classification capability. This demonstrator's best performance was in detection of clusters, with $P_{match} = 0.33$. This demonstrator's worst performance was in detecting mortars, with $P_{match} = 0.05$. There were no mines in the area searched by this demonstrator. The performance of this demonstrator was best for deep targets. When the

targets are grouped as those buried less than 3 feet, between 3 and 6 feet, and more than 6 feet deep, P_{match} was 0.13, 0.05, and 0.27, respectively. The demonstrator was best able to detect large size targets, with $P_{match} = 0.21$, and detected small and medium targets at $P_{match} = 0.07$ and $P_{match} = 0.10$, respectively.

The detection capabilities calculated above are all based on P_{match}, the probability of a one-to-one match between a demonstrator declaration and an emplaced target. This method of determining detection probability assumes a high degree of sensor resolution, which may not be justified if a sensor cannot resolve signals resulting from two (or more) closely spaced objects. Two alternative measures of detection capability are also considered. For demonstrator declarations with multiple baseline items within R_{crit}, P_{group} credits the demonstrator with detecting a single group of targets, and does not penalize for any missed detections. P_{near} measures the probability that an item will be found if holes of radius R_{crit} are dug on all demonstrator declarations, allowing multiple baseline targets to be matched to a single demonstrator declaration. The following table compares detection capability calculated using these three measures, where P_{group} and P_{match} represent upper and lower bounds on the probability of detection, P_d, achieved on this demonstration. The false alarm rate, *FAR*, will be the same for all three measures; the false negative ratio will be different. Also presented in the table are P_{match} and P_{near} calculated with the plastic mines removed from the baseline. All detection measures are for ordnance items only.

Calculation method	Value
P_{match}[1]	0.13
P_{match}(no mines)	0.13
P_{group}	0.13
P_{near}	0.11
P_{near} (no mines)	0.11

[1] For this demonstration, the correction to detection capabilities for random matches was small (< 0.03).

The following tables provide information about specific demonstrator declarations and baseline items which may be of use to the demonstrator in assessing target identification capabilities for future operations, or for making improvements in the sensor or the data processing algorithms. The correct locations and types are provided for 17 baseline items that were dug up during or shortly after the remediation demonstrations. Five of these baseline items are isolated, individual objects. The remaining 12 are six closely spaced pairs of items. These items have either been removed from the site or were

emplaced again at new locations. Three demonstrator declarations that were not within 2 meters of a baseline item are also provided.

Locations and Types of Selected Baseline Targets				
Type	Size	X[a]	Y	Depth
Bomb	Large	2104200.6	14139531.2	8.0
Bomb	Large	2105272.6	14139550.2	5.6
Bomb	Large	2105272.5	14139555.1	2.9
Projectile	Medium	2105087.4	14139006.9	3.4
Projectile	Medium	2105105.8	14139336.8	2.7
Projectile	Small	2105134.5	14139561.0	6.1
Projectile	Medium	2105326.1	14139555.1	2.0
Projectile	Medium	2105014.3	14139251.5	3.3
Mortar	Small	2105139.5	14139036.9	3.1
Mortar	Small	2105327.9	14139554.6	1.2
Mortar	Small	2104978.4	14139240.2	1.3
Mortar	Small	2104976.0	14139229.1	0.8
Mortar	Small	2104976.7	14139229.3	0.9
Mortar	Small	2104013.6	14138535.2	1.1
Mortar	Small	2104017.5	14138533.9	0.9
NonOrd	Large	2105014.7	14139251.1	0.9
NonOrd	Large	2105139.8	14139037.1	1.1
[a] All dimensions are in feet.				

Demonstrator Declarations Not Within 2 m of a Baseline Item	
110[a]	No known baseline item within 2 m of this location.
301	No known baseline item within 2 m of this location.
410	No known baseline item within 2 m of this location.
[a] Target numbers are from the demonstrator numbering scheme.	

III. SUMMARY

The purpose of the 1994 Jefferson Proving Ground demonstration was to evaluate the state of the art in UXO detection and remediation technology. For all detection demonstrators, the maximum detection capabilities as determined by the target matching algorithm

have been lower than $P_{match} = 0.65$ and there have been multiple false declarations per ordnance item detected. Further, statistical analysis of the JPG data indicates that demonstrators with similar detection capabilities cannot be ordinally ranked with confidence. Therefore, demonstrators are binned on the basis of P_{match}, false alarm rate, and a separation criteria d, which combines measures of P_{match} and P_{fa} into a single variable and allows for comparison of demonstrators with widely different detection capabilities and false alarm rates. In this assessment, Chemrad/EG&G performed significantly worse than the average JPG demonstrator in terms of P_{match}, in the third of four bins in false alarms, and in the last of four bins in the separation measure.

This page is intentionally left blank

EVALUATION OF CHEMRAD CORP. (G-822L) 40-ACRE SITE

I. DEMONSTRATOR RESULTS

	Entire Site	Area Searched[1]
Detection Capability		
overall detection ratio, R_{all}	0.28	0.28
ordnance detection ratio, R_{ord}	0.26	0.26
nonordnance detection ratio, R_{nonord}	0.33	0.33
mistyped ordnance ratio, *MR*	1.0‡	1.0‡
False Positive Ratio, *FPR*	0.0‡	0.0‡
False Negative Ratio, *FNR*	*‡	*‡
False Alarm Rate, *FAR*	1.9	1.9
Radial Accuracy, ΔR_{xy} *(feet)*	3.8	3.8
Depth Accuracy, ΔR_z *(feet)*	2.5	2.5
Target Classification Capabilities		
single targets	0.0	0.0
multiple targets	0.0	0.0
bombs	0.0	0.0
projectiles	0.0	0.0
mortars	0.0	0.0
mines	*	*
clusters	0.0	0.0

1 These columns are identical because this demonstrator searched the entire 40-acre site.

* Ratio is undefined (i.e., zero denominator). The false negative ratio is undefined because all targets are identified as an ordnance. The classification ratios are undefined because no targets of the class were located.

‡ The values shown here are a consequence of the demonstrator declaring all items as nonordnance.

Survey Rate

The demonstrator searched the entire 40-acre site within the allotted 1 week. The proposal gave survey rates as 10 acres per day for hand-held magnetometer and 20–40 acres per day for the surface-towed magnetometer.

II. DISCUSSION OF DEMONSTRATOR PERFORMANCE

Chemrad used an 822L magnetometer, with a range of 20–25 feet for detection, and a USRADS acoustic positioning system for navigation.

Because all detections by this demonstrator were classified as nonordnance items, several of the ratios used in this evaluation are not useful for characterizing this demonstrator. Most seriously, the 0.0 value for the false positive ratio implies that this demonstrator has no nonordnance items falsely declared as ordnance. However, this number indicates only that no items were so declared. The false negative ratio is undefined because its denominator requires that some items be declared as ordnance; a more accurate measure of this demonstrator's false negative ratio would include items declared as nonordnance as well. *MR* has a value of 1.0 as all ordnance was mistyped (i.e., declared nonordnance). Also, all classification ratios are zero because the demonstrator did not attempt to classify its declarations by ordnance type, with the exception of the mine classification ratio, which is undefined as the demonstrator detected no mines.

Corrections for random hits by this demonstrator reduced the uncorrected P_{match} = 0.26 to a corrected P_{match} = 0.24, where P_{match} refers to the ordnance detection capability.

Detection capabilities were calculated for the various ordnance types emplaced. Credit is given for a detection regardless of classification ability. This demonstrator's best performance was in the detection of bombs, with P_{match} = 0.74. This demonstrator's worst performance was in detecting mines, with zero detections of this target type. Since the sensor used by this demonstrator is only able to detect metal objects, the inability to detect plastic mines is to be expected. This demonstrator's performance in the detection of mortars was also poor, with P_{match} = 0.07. The performance of this demonstrator was best at intermediate and deep depths. When the targets are grouped as those buried less than 3 feet, between 3 and 6 feet, and more than 6 feet deep, P_{match} was 0.20, 0.37, and 0.60, respectively. The demonstrator was best able to detect large targets, with P_{match} = 0.52, and detected small and medium targets at P_{match} = 0.42 and P_{match} = 0.08, respectively.

The detection capabilities calculated above are all based on P_{match}, the probability of a one-to-one match between a demonstrator declaration and an emplaced target. This method of determining detection probability assumes a high degree of sensor resolution, which may not be justified if a sensor cannot resolve signals resulting from two (or more) closely spaced objects. Two alternative measures of detection capability are also considered. For demonstrator declarations with multiple baseline items within R_{crit}, P_{group} credits the demonstrator with detecting a single group of targets, and does not penalize for any missed detections. P_{near} measures the probability that an item will be found if holes of radius R_{crit} are dug on all demonstrator declarations, allowing multiple baseline targets to be matched to a single demonstrator declaration. The following table compares detection capability calculated using these three measures, where P_{group} and P_{match} represent upper and lower bounds on the probability of detection, P_d, achieved on this demonstration. The false alarm rate, *FAR*, will be the same for all three measures; the false negative ratio will be different. Also presented in the table are P_{match} and P_{near} calculated with the plastic mines removed from the baseline. All detection measures are for ordnance items only.

Calculation method	Value
P_{match}[1]	0.26
P_{match}*(no mines)*	0.30
P_{group}	0.28
P_{near}	0.33
P_{near} *(no mines)*	0.38

1 For this demonstration, the correction to detection capabilities for random matches was small (< 0.03).

The following tables provide information about specific demonstrator declarations and baseline items which may be of use to the demonstrator in assessing target identification capabilities for future operations, or for making improvements in the sensor or the data processing algorithms. The correct locations and types are provided for 17 baseline items that were dug up during or shortly after the remediation demonstrations. Five of these baseline items are isolated, individual objects. The remaining 12 are six closely spaced pairs of items. These items have either been removed from the site or were emplaced again at new locations. Three demonstrator declarations that were not within 2 meters of a baseline item are also provided.

Locations and Types of Selected Baseline Targets				
Type	Size	X[a]	Y	Depth
Bomb	Large	2104200.6	14139531.2	8.0
Bomb	Large	2105272.6	14139550.2	5.6
Bomb	Large	2105272.5	14139555.1	2.9
Projectile	Medium	2105087.4	14139006.9	3.4
Projectile	Medium	2105105.8	14139336.8	2.7
Projectile	Small	2105134.5	14139561.0	6.1
Projectile	Medium	2105326.1	14139555.1	2.0
Projectile	Medium	2105014.3	14139251.5	3.3
Mortar	Small	2105139.5	14139036.9	3.1
Mortar	Small	2105327.9	14139554.6	1.2
Mortar	Small	2104978.4	14139240.2	1.3
Mortar	Small	2104976.0	14139229.1	0.8
Mortar	Small	2104976.7	14139229.3	0.9
Mortar	Small	2104013.6	14138535.2	1.1
Mortar	Small	2104017.5	14138533.9	0.9
NonOrd	Large	2105014.7	14139251.1	0.9
NonOrd	Large	2105139.8	14139037.1	1.1
[a] All dimensions are in feet.				

Demonstrator Declarations Not Within 2 m of a Baseline Item	
50[a]	No known baseline item within 2 m of this location.
90	No known baseline item within 2 m of this location.
180	No known baseline item within 2 m of this location.
[a] Target numbers are from the demonstrator numbering scheme.	

III. SUMMARY

The purpose of the 1994 Jefferson Proving Ground demonstration was to evaluate the state of the art in UXO detection and remediation technology. For all detection demonstrators, the maximum detection capabilities as measured by the target matching algorithm have been lower than $P_{match} = 0.65$ and there have been multiple false declarations per ordnance item detected. Further, statistical analysis of the JPG data indicates that

demonstrators with similar detection capabilities cannot be ordinally ranked with confidence. Therefore, demonstrators are binned on the basis of P_{match}, false alarm rate, and a separation criteria d, which combines measures of P_{match} and P_{fa} into a single variable and allows for comparison of demonstrators with widely different detection capabilities and false alarm rates. In this assessment, Chemrad (G-822L) performed in the vicinity of the average JPG demonstrator in terms of P_{match}, in the second of four bins in false alarms, and in the third of four bins in the separation measure.

This page is intentionally left blank

EVALUATION OF CHEMRAD (GSM-19)
40-ACRE SITE

I. DEMONSTRATOR RESULTS

	Entire Site	Area Searched[1]
Detection Capability		
overall detection ratio, R_{all}	0.05	0.05
ordnance detection ratio, R_{ord}	0.04	0.04
nonordnance detection ratio, R_{nonord}	0.07	0.07
mistyped ordnance ratio, MR	0.0‡	0.0‡
False Positive Ratio, FPR	1.00‡	1.00‡
False Negative Ratio, FNR	0.97	0.97
False Alarm Rate, FAR	1.5	1.5
Radial Accuracy, ΔR_{xy} *(feet)*	5.19	5.19
Depth Accuracy, ΔR_z *(feet)*	1.92	1.92
Target Classification Capabilities		
single targets	1.0	1.0
multiple targets	*	*
bombs	0.0	0.0
projectiles	0.0	0.0
mortars	0.0	0.0
mines	*	*
clusters	*	*

[1] These columns are identical because this demonstrator searched the entire 40-acre site.

* Ratio is undefined (i.e., zero denominator). A classification ratio is undefined if no targets of that class are detected by the demonstrator.

‡ The values shown here are a consequence of the demonstrator declaring all items as ordnance.

Survey Rate

The demonstrator searched the 40-acre site within the allotted 1 week. The proposal gave the survey rate as 40 acres per 4 days.

II. DISCUSSION OF DEMONSTRATOR PERFORMANCE

Chemrad used a GSM-19 magnetometer/gradiometer to survey the site and an acoustic USRADS surveying tool for navigation. The accuracy of the navigation tool is 6 inches. The sensor is not able to determine the orientation of a projectile in the ground.

This demonstrator did not attempt to determine whether items were ordnance or nonordnance, typing all detections as ordnance items. Because no attempt was made to distinguish ordnance from nonordnance, several of the ratios used in this evaluation are not useful for characterizing this demonstrator. For example, $MR = 0$ not because the demonstrator made no errors in separating ordnance from nonordnance; rather, because no items were typed as nonordnance, no ordnance items could be misidentified as non-ordnance. $FPR = 1.0$ since any nonordnance items found were typed as ordnance. Of particular note for this demonstrator is the false negative ratio of 0.97. As a practical matter, if all the ordnance declarations of this demonstrator were investigated, some holes would contain emplaced nonordnance items, and 97% of the remaining holes dug for remediation would not contain any of the ordnance items known by the government to be present.

Corrections for random hits by this demonstrator reduced the uncorrected P_{match} = 0.04 to a corrected $P_{match} = 0.02$, where P_{match} refers to the ordnance detection capability.

Detection capabilities were calculated for the various ordnance types emplaced. Credit was given for a detection regardless of classification ability, when calculating the detection capabilities reported in this paragraph. This demonstrator's best performance was in the detection of bombs, with $P_{match} = 0.16$. This demonstrator's worst performance was in the detection of mines and cluster targets, with zero detections of either target type. Since the sensors used by this demonstrator can detect only metal objects, the inability to detect plastic mines is to be expected. The performance of this demonstrator was best for targets at medium depths. When the targets are grouped as those buried less than 3 feet, between 3 and 6 feet, and more than 6 feet deep, P_{match} was 0.04, 0.09, and 0.04, respectively. The demonstrator was best able to detect large and medium size targets, with $P_{match} = 0.07$ and $P_{match} = 0.06$, respectively, and detected small targets at $P_{match} = 0.03$.

The detection capabilities calculated above are all based on P_{match}, the probability of a one-to-one match between a demonstrator declaration and an emplaced target. This method of determining detection probability assumes a high degree of sensor resolution, which may not be justified if a sensor cannot resolve signals resulting from two (or more) closely spaced objects. Two alternative measures of detection capability are also considered. For demonstrator declarations with multiple baseline items within R_{crit}, P_{group} credits the demonstrator with detecting a single group of targets, and does not penalize for any missed detections. P_{near} measures the probability that an item will be found if holes of radius R_{crit} are dug on all demonstrator declarations, allowing multiple baseline targets to be matched to a single demonstrator declaration. The following table compares detection capability calculated using these three measures, where P_{group} and P_{match} represent upper and lower bounds on the probability of detection, P_d, achieved on this demonstration. The false alarm rate, *FAR*, will be the same for all three measures; the false negative ratio will be different. Also presented in the table are P_{match} and P_{near} calculated with the plastic mines removed from the baseline. All detection measures are for ordnance items only.

Calculation method	Value
P_{match}[1]	0.04
P_{match} (no mines)	0.05
P_{group}	0.04
P_{near}	0.04
P_{near} (no mines)	0.04

[1] For this demonstration, the correction to detection capabilities for random matches was small (< 0.03).

The following tables provide information about specific demonstrator declarations and baseline items which may be of use to the demonstrator in assessing target identification capabilities for future operations, or for making improvements in the sensor or the data processing algorithms. The correct locations and types are provided for 17 baseline items that were dug up during or shortly after the remediation demonstrations. Five of these baseline items are isolated, individual objects. The remaining 12 are six closely spaced pairs of items. These items have either been removed from the site or were emplaced again at new locations. Three demonstrator declarations that were not within 2 meters of a baseline item are also provided.

Locations and Types of Selected Baseline Targets				
Type	Size	X[a]	Y	Depth
Bomb	Large	2104200.6	14139531.2	8.0
Bomb	Large	2105272.6	14139550.2	5.6
Bomb	Large	2105272.5	14139555.1	2.9
Projectile	Medium	2105087.4	14139006.9	3.4
Projectile	Medium	2105105.8	14139336.8	2.7
Projectile	Small	2105134.5	14139561.0	6.1
Projectile	Medium	2105326.1	14139555.1	2.0
Projectile	Medium	2105014.3	14139251.5	3.3
Mortar	Small	2105139.5	14139036.9	3.1
Mortar	Small	2105327.9	14139554.6	1.2
Mortar	Small	2104978.4	14139240.2	1.3
Mortar	Small	2104976.0	14139229.1	0.8
Mortar	Small	2104976.7	14139229.3	0.9
Mortar	Small	2104013.6	14138535.2	1.1
Mortar	Small	2104017.5	14138533.9	0.9
NonOrd	Large	2105014.7	14139251.1	0.9
NonOrd	Large	2105139.8	14139037.1	1.1

[a] All dimensions are in feet.

Demonstrator Declarations Not Within 2 m of a Baseline Item	
10[a]	No known baseline item within 2 m of this location.
80	No known baseline item within 2 m of this location.
200	No known baseline item within 2 m of this location.

[a] Target numbers are from the demonstrator numbering scheme.

III. SUMMARY

The purpose of the 1994 Jefferson Proving Ground demonstration was to evaluate the state of the art in UXO detection and remediation technology. For all detection demonstrators, the maximum detection capabilities as measured by the target matching algorithm have been lower than $P_{match} = 0.65$ and there have been multiple false declarations per ordnance item detected. Further, statistical analysis of the JPG data indicates that demon-

strators with similar detection capabilities cannot be ordinally ranked with confidence. Therefore, demonstrators are binned on the basis of P_{match}, false alarm rate, and a separation criteria d, which combines measures of P_{match} and P_{fa} into a single variable and allows for comparison of demonstrators with widely different detection capabilities and false alarm rates. In this assessment, Chemrad (GSM-19) performed significantly worse than the average JPG demonstrator in terms of P_{match}, in the second of four bins in false alarms, and in the last bin of four bins in the separation measure.

This page is intentionally left blank

EVALUATION OF COLEMAN RESEARCH CORPORATION 40-ACRE SITE

I. DEMONSTRATOR RESULTS

	Entire Site	Area Searched[1]
Detection Capability		
overall detection ratio, R_{all}	0.33	0.33
ordnance detection ratio, R_{ord}	0.39	0.39
nonordnance detection ratio, R_{nonord}	0.20	0.20
mistyped ordnance ratio, *MR*	0.0‡	0.0‡
False Positive Ratio, *FPR*	1.0‡	1.0‡
False Negative Ratio, *FNR*	0.93	0.93
False Alarm Rate, *FAR*	5.0	5.2
Radial Accuracy, ΔR_{xy} *(feet)*	4.0	3.98
Depth Accuracy, ΔR_z *(feet)*	1.5	1.5
Target Classification Capabilities		
single targets	0.98	0.98
multiple targets	0.0	0.0
bombs	1.00	1.00
projectiles	0.27	0.303
mortars	0.0	0.0
mines	0.0	0.0
clusters	0.0	0.0

[1] The portion of the site searched by this demonstrator may not be representative of the overall difficulty of the entire 40-acre site.

‡ The values shown here are a consequence of the demonstrator declaring all items as ordnance.

Survey Rate

The demonstrator searched approximately 36 acres within the allotted 1 week. The system suffered several minor mechanical failures, such as flat tires and broken hinges,

which resulted in a total of about 9 hours of down-time during the course of the week. The proposal gave survey rates as 40 acres in 4 days.

II. DISCUSSION OF DEMONSTRATOR PERFORMANCE

Coleman Research Corporation used a towed multisensor array system (TOMAS) which included both a ground penetrating radar and an induction coil detector. A differential Global Positioning System was used for navigation. CRC proposed to use several radar technologies in the demonstration, including an Earth Penetrating Radar Imaging System, a hand-held radar mine detector, an advanced system for close proximity robotic mine detection (which fuses infrared and radar sensors), and a radar system designed for locating nonmetallic objects. The proposal states that the sensors will work to a depth of 7 meters. **No mention is made in the proposal of induction coil sensors.**

This demonstrator did not attempt to determine whether items were ordnance or nonordnance, typing all detections as ordnance items. Because no attempt was made to distinguish ordnance from nonordnance, several of the ratios used in this evaluation are not useful for characterizing this demonstrator. For example, $MR = 0$ not because the demonstrator made no errors in separating ordnance from nonordnance; rather, because no items were typed as nonordnance, no ordnance items could be misidentified as nonordnance. $FPR = 1.0$ since any nonordnance items found were typed as ordnance.

This demonstrator searched approximately 36 acres of the 40-acre site. When the demonstrator's detection capability and false alarm rates are recalculated to include only the portion of the site searched, the results are as follows:

$$P_{match} = 0.33$$

$$P_{match}(\text{ord}) = 0.39$$

$$FAR = 5.2$$

Corrections for random hits by this demonstrator reduced the uncorrected $P_{match} = 0.39$ to a corrected $P_{match} = 0.34$, where P_{match} refers to the ordnance detection capability on only the area searched.

Detections were specified by the demonstrator as having High, Medium and Low probability of containing a target item. When the evaluation ratios are recalculated to include only the High probability detections, and both the High and Medium probability detections, the results are as follows:

	High	High + Medium
P_{match}	0.26	0.37
$P_{match(corr)}$	0.26	0.35
FAR	0.57	2.2
FNR	0.68	0.85

P_{match} here considers only the ordnance items. Eliminating the Low confidence detections does not change the demonstrator's P_{match} substantially, but decreases the false alarm rate by more than half. For the measures of demonstrator performance that follow, the low confidence targets were removed from the demonstrator's target declarations.

Detection capabilities were calculated for the various ordnance types emplaced. Only the targets in the area searched were considered, and credit was given for a detection regardless of classification ability, when calculating the detection capabilities reported in this paragraph. This demonstrator's best performance was in detection of projectiles, with P_{match} = 0.55. Detection ability for bombs and mortars were P_{match} = 0.42 and 0.32, respectively. This demonstrator's worst performance was in detecting mines, with zero detections. The performance of this demonstrator was best at intermediate depths. When the targets are grouped as those buried less than 3 feet, between 3 and 6 feet, and more than 6 feet deep, P_{match} was 0.28, 0.41, and 0.32, respectively. The demonstrator was best able to detect medium size targets, with P_{match} = 0.53, and detected small and large targets at P_{match} = 0.25 and P_{match} = 0.27, respectively.

The detection capabilities calculated above are all based on P_{match}, the probability of a one-to-one match between a demonstrator declaration and an emplaced target. This method of determining detection probability assumes a high degree of sensor resolution, which may not be justified if a sensor cannot resolve signals resulting from two (or more) closely spaced objects. Two alternative measures of detection capability are also considered. For demonstrator declarations with multiple baseline items within R_{crit}, P_{group} credits the demonstrator with detecting a single group of targets, and does not penalize for any missed detections. P_{near} measures the probability that an item will be found if holes of radius R_{crit} are dug on all demonstrator declarations, allowing multiple baseline targets to be matched to a single demonstrator declaration. The following table compares detection capability calculated using these three measures, where P_{group} and P_{match} represent upper and lower bounds on the probability of detection, P_d, achieved on this demonstration. The false alarm rate, FAR, will be the same for all three measures; the false negative ratio will

be different. Also presented in the table are P_{match} and P_{near} calculated with the plastic mines removed from the baseline. All detection measures are for ordnance items only.

Calculation method	Value
P_{match}[1]	0.37
P_{match}(corrected)	0.42
P_{group}	0.40
P_{near}	0.44
P_{near} (no mines)	0.50

1 For this demonstration, the correction to detection capabilities for random matches was small (< 0.03).

The following tables provide information about specific demonstrator declarations and baseline items which may be of use to the demonstrator in assessing target identification capabilities for future operations, or for making improvements in the sensor or the data processing algorithms. The correct locations and types are provided for 17 baseline items that were dug up during or shortly after the remediation demonstrations. Five of these baseline items are isolated, individual objects. The remaining 12 are six closely spaced pairs of items. These items have either been removed from the site or were emplaced again at new locations. Three demonstrator declarations that were not within 2 meters of a baseline item are also provided.

Locations and Types of Selected Baseline Targets				
Type	Size	X[a]	Y	Depth
Bomb	Large	2104200.6	14139531.2	8.0
Bomb	Large	2105272.6	14139550.2	5.6
Bomb	Large	2105272.5	14139555.1	2.9
Projectile	Medium	2105087.4	14139006.9	3.4
Projectile	Medium	2105105.8	14139336.8	2.7
Projectile	Small	2105134.5	14139561.0	6.1
Projectile	Medium	2105326.1	14139555.1	2.0
Projectile	Medium	2105014.3	14139251.5	3.3
Mortar	Small	2105139.5	14139036.9	3.1
Mortar	Small	2105327.9	14139554.6	1.2
Mortar	Small	2104978.4	14139240.2	1.3
Mortar	Small	2104976.0	14139229.1	0.8
Mortar	Small	2104976.7	14139229.3	0.9
Mortar	Small	2104013.6	14138535.2	1.1
Mortar	Small	2104017.5	14138533.9	0.9
NonOrd	Large	2105014.7	14139251.1	0.9
NonOrd	Large	2105139.8	14139037.1	1.1
[a] All dimensions are in feet.				

Demonstrator Declarations Not Within 2 m of a Baseline Item	
11[a]	No known baseline item within 2 m of this location.
33	No known baseline item within 2 m of this location.
948	No known baseline item within 2 m of this location.
[a] Target numbers are from the demonstrator numbering scheme.	

III. SUMMARY

The purpose of the 1994 Jefferson Proving Ground demonstration was to evaluate the state of the art in UXO detection and remediation technology. For all detection demonstrators, the maximum detection capabilities as measured by the target matching algorithm have been lower than $P_{match} = 0.65$ and there have been multiple false declarations per ordnance item detected. Further, statistical analysis of the JPG data indicates that

demonstrators with similar detection capabilities cannot be ordinally ranked with confidence. Therefore, demonstrators are binned on the basis of P_{match}, false alarm rate, and a separation criteria d, which combines measures of P_{match} and P_{fa} into a single variable and allows for comparison of demonstrators with widely different detection capabilities and false alarm rates. In this assessment, Coleman performed significantly better than the average JPG demonstrator in terms of P_{match}, in the second of four bins in false alarms, and in the second of four bins in the separation measure, when the low confidence declarations are removed from the demonstrator data set.

EVALUATION OF DYNAMIC SYSTEMS, INC.
40-ACRE SITE

I. DEMONSTRATOR RESULTS

	Entire Site	Area Searched[1]
Detection Capability		
overall detection ratio, R_{all}	0.05	0.34
ordnance detection ratio, R_{ord}	0.04	0.25
nonordnance detection ratio, R_{nonord}	0.07	0.63
mistyped ordnance ratio, MR	0.0‡	0.0‡
False Positive Ratio, FPR	1.00‡	1.00‡
False Negative Ratio, FNR	0.68	0.67
False Alarm Rate, FAR	0.07	0.5
Radial Accuracy, ΔR_{xy} *(feet)*	1.47	1.47
Depth Accuracy, ΔR_z *(feet)*	1.30	1.30
Target Classification Capabilities		
single targets	1	1
multiple targets	*	*
bombs	1	1
projectiles	1	1
mortars	0	0
mines	*	*
clusters	*	*

1 The portion of the site searched by this demonstrator may not be representative of the overall difficulty of the entire 40-acre site.

* Ratio is undefined (i.e., zero denominator). A classification ratio is undefined if no targets of that class are located by the demonstrator.

‡ The values shown here are a consequence of the demonstrator declaring all items as ordnance.

Survey Rate

The demonstrator searched approximately 5.5 acres of the 40-acre site within the allotted one week. The proposal gave survey rates as 10 acres per day for hand-held magnetometer and 20–40 acres per day for the surface towed magnetometer. Due to a misunderstanding, the system was not brought on to the site until 1030 hours on the first day of the demonstration. The data from 1 day's surveying was corrupted due to a disconnected cable that was not discovered until evening. However, the demonstrator report did not indicate the reason the survey rate fell so far short of the demonstrator's initial expectations.

II. DISCUSSION OF DEMONSTRATOR PERFORMANCE

Dynamic Systems used a man-portable system consisting of a Billigsley magnetometer and a Foerster magnetometer for detection, and a TopCon 302 Survey instrument for navigation. Both sensors had a range of 15 to 20 feet. The demonstrator report notes that the surface was rough with new vegetation.

This demonstrator did not attempt to determine whether items were ordnance or nonordnance, typing all detections as ordnance items. Because no attempt was made to distinguish ordnance from nonordnance, several of the ratios used in this evaluation are not useful for characterizing this demonstrator. For example, $MR = 0$ not because the demonstrator made no errors in separating ordnance from nonordnance; rather, because no items were typed as nonordnance, no ordnance items could be misidentified as nonordnance. $FPR = 1.0$ since any nonordnance items found were typed as ordnance.

The demonstrator surveyed approximately 5.5 acres of the 40-acre site. When the demonstrator's detection capability and false alarm rates are recalculated to include only the portion of the site searched, the results are as follows:

$$P_{match} = 0.34$$

$$P_{match}(\text{ord}) = 0.25$$

$$FAR = 0.5$$

Corrections for random hits by this demonstrator reduced the uncorrected $P_{match} = 0.25$ to a corrected $P_{match} = 0.24$, where P_{match} refers to ordnance items on only the area searched.

Detection capabilities were calculated for the various ordnance types emplaced. Only the targets in the area searched were considered and credit was given for a detection regardless of classification ability, when calculating the detection capabilities reported in this paragraph. This demonstrator's best performance was in detection of bombs, with $P_{match} = 0.50$. This demonstrator's worst performance was in detecting clusters, with zero detections of this target type. The region searched by this demonstrator did not contain any mines, which have presented the greatest detection challenge to other demonstrators. The performance of this demonstrator was best in detection of intermediate depth targets. When the targets are grouped as those buried less than 3 feet, between 3 and 6 feet, and more than 6 feet deep, P_{match} was 0.33, 0.43, and 0.25, respectively. The demonstrator was best able to detect large targets, with $P_{match} = 0.45$, and detected small and medium targets at $P_{match} = 0.31$ and $P_{match} = 0.25$, respectively.

The detection capabilities calculated above are all based on P_{match}, the probability of a one-to-one match between a demonstrator declaration and an emplaced target. This method of determining detection probability assumes a high degree of sensor resolution, which may not be justified if a sensor cannot resolve signals resulting from two (or more) closely spaced objects. Two alternative measures of detection capability are also considered. For demonstrator declarations with multiple baseline items within R_{crit}, P_{group} credits the demonstrator with detecting a single group of targets, and does not penalize for any missed detections. P_{near} measures the probability that an item will be found if holes of radius R_{crit} are dug on all demonstrator declarations, allowing multiple baseline targets to be matched to a single demonstrator declaration. The following table compares detection capability calculated using these three measures, where P_{group} and P_{match} represent upper and lower bounds on the probability of detection, P_d, achieved on this demonstration. The false alarm rate, *FAR*, will be the same for all three measures; the false negative ratio will be different. Also presented in the table are P_{match} and P_{near} calculated with the plastic mines removed from the baseline. All detection measures are for ordnance items only.

Calculation method	Value
P_{match}[1]	0.25
P_{match}(no mines)	0.25
P_{group}	0.32
P_{near}	0.46
P_{near} (no mines)	0.46

[1] For this demonstration, the correction to detection capabilities for random matches was small (< 0.03).

The following tables provide information about specific demonstrator declarations and baseline items which may be of use to the demonstrator in assessing target identification capabilities for future operations, or for making improvements in the sensor or the data processing algorithms. The correct locations and types are provided for 17 baseline items that were dug up during or shortly after the remediation demonstrations. Five of these baseline items are isolated, individual objects. The remaining 12 are six closely spaced pairs of items. These items have either been removed from the site or were emplaced again at new locations. Three demonstrator declarations that were not within 2 meters of a baseline item are also provided.

Locations and Types of Selected Baseline Targets				
Type	Size	X[a]	Y	Depth
Bomb	Large	2104200.6	14139531.2	8.0
Bomb	Large	2105272.6	14139550.2	5.6
Bomb	Large	2105272.5	14139555.1	2.9
Projectile	Medium	2105087.4	14139006.9	3.4
Projectile	Medium	2105105.8	14139336.8	2.7
Projectile	Small	2105134.5	14139561.0	6.1
Projectile	Medium	2105326.1	14139555.1	2.0
Projectile	Medium	2105014.3	14139251.5	3.3
Mortar	Small	2105139.5	14139036.9	3.1
Mortar	Small	2105327.9	14139554.6	1.2
Mortar	Small	2104978.4	14139240.2	1.3
Mortar	Small	2104976.0	14139229.1	0.8
Mortar	Small	2104976.7	14139229.3	0.9
Mortar	Small	2104013.6	14138535.2	1.1
Mortar	Small	2104017.5	14138533.9	0.9
NonOrd	Large	2105014.7	14139251.1	0.9
NonOrd	Large	2105139.8	14139037.1	1.1

[a] All dimensions are in feet.

Demonstrator Declarations Not Within 2 m of a Baseline Item	
101[a]	No known baseline item within 2 m of this location.
117	No known baseline item within 2 m of this location.
122	No known baseline item within 2 m of this location.

[a] Target numbers are from the demonstrator numbering scheme.

III. SUMMARY

The purpose of the 1994 Jefferson Proving Ground demonstration was to evaluate the state of the art in UXO detection and remediation technology. For all detection demonstrators, the maximum detection capabilities as measured by the target matching algorithm have been lower than $P_{match} = 0.65$ and there have been multiple false declarations per ordnance item detected. Further, statistical analysis of the JPG data indicates that demonstrators with similar detection capabilities cannot be ordinally ranked with confidence. Therefore, demonstrators are binned on the basis of P_{match}, false alarm rate, and a separation criteria d, which combines measures of P_{match} and P_{fa} into a single variable and allows for comparison of demonstrators with widely different detection capabilities and false alarm rates. In this assessment, Dynamic performed in the vicinity of the average JPG demonstrator in terms of P_{match}, in the top bin of four bins in false alarms, and in the second of four bins in the separation measure.

This page is intentionally left blank

EVALUATION OF ENSCO
40-ACRE SITE

I. DEMONSTRATOR RESULTS

	Entire Site	Area Searched[1]
Detection Capability		
overall detection ratio, R_{all}	0.004	0.03
ordnance detection ratio, R_{ord}	0.006	0.04
nonordnance detection ratio, R_{nonord}	0.0	0.0
mistyped ordnance ratio, *MR*	0.0‡	0.0‡
False Positive Ratio, *FPR*	*	*
False Negative Ratio, *FNR*	1.0	1.0
False Alarm Rate, *FAR*	1.28	4.8
Radial Accuracy, ΔR_{xy} *(feet)*	6.5	6.5
Depth Accuracy, ΔR_z *(feet)*	0.5	0.5
Target Classification Capabilities		
single targets	1.0	1.0
multiple targets	*	*
bombs	*	*
projectiles	0	0
mortars	*	*
mines	*	*
clusters	*	*

[1] The portion of the site searched by this demonstrator may not be representative of the overall difficulty of the entire 40-acre site.

* Ratio is undefined (i.e., zero denominator). A classification ratio is undefined if no targets of that class are located by the demonstrator.

‡ The values shown here are a consequence of the demonstrator declaring all items as ordnance.

Survey Rate

The demonstrator searched approximately 10 acres of the 40-acre site within the allotted one week. The proposal implied that the entire site could be surveyed in the allotted time search in a "continuous" mode The demonstrator report does not indicate the reason why the survey rate fell short of that initially expected by the demonstrator.

II. DISCUSSION OF DEMONSTRATOR PERFORMANCE

ENSCO used several ground penetrating radars towed on a sled pulled by a modified golf cart for detection. A survey wheel and laser Track were used for navigation.

This demonstrator did not attempt to determine whether items were ordnance or nonordnance, typing all detections as ordnance items. Because no attempt was made to distinguish ordnance from nonordnance, several of the ratios used in this evaluation are not useful for characterizing this demonstrator. For example, MR = 0 not because the demonstrator made no errors in separating ordnance from nonordnance; rather, because no items were typed as nonordnance, no ordnance items could be misidentified as nonordnance. The importance of the 1.0 value for the false negative ratio should not be overlooked. As a practical matter, if all the ordnance declarations of this demonstrator were explored for remediation, some holes dug would contain emplaced nonordnance items and the remaining holes dug would not contain any of the ordnance items known by the government to be present.

The demonstrator surveyed approximately 10 acres of the 40-acre site. When the demonstrator's detection capability and false alarm rates are recalculated to include only the portion of the site searched, the results are as follows:

$$P_{match} = 0.03$$

$$P_{match}(\text{ord}) = 0.04$$

$$FAR = 4.8$$

Corrections for random hits by this demonstrator reduced the uncorrected P_{match} = 0.04 to a corrected P_{match} = – 0.03, where P_{match} refers to the ordnance detection capability on only the area searched.

The detection capability of this demonstrator showed a strong dependence on the choice of critical radius. At R_{crit} = 2 m, the P_{match} (ord) is 0.04, but when R_{crit} is increased to 5 m, P_{match} (ord) grows to 0.29. This is a much larger increase than is seen

for any other demonstrator, and could arise from this demonstrator having poorer locating capabilities. As R_{crit} increases, the number of fortuitous matches between demonstrator false alarms and baseline targets will also increase. To account for this effect, P_{match} observed and P_{match} corrected were calculated for increasing R_{crit} at 2-foot intervals. The corrected P_{match} is flat in the region of 10–12 feet (3.0–3.7 m), indicating that most true detections are captured within this range of R_{crit}. This analysis gives a corrected P_{match} = 0.11, where the observed uncorrected values of P_{match} were 0.25 and 0.29 for 10 feet and 12 feet, respectively.

This demonstrator found so few ordnance items for the 2-m radius measurements that the classification ratios and measures of detection capability in terms of target class, size, and depth were meaningless.

The detection capabilities calculated above are all based on P_{match}, the probability of a one-to-one match between a demonstrator declaration and an emplaced target. This method of determining detection probability assumes a high degree of sensor resolution, which may not be justified if a sensor cannot resolve signals resulting from two (or more) closely spaced objects. Two alternative measures of detection capability are also considered. For demonstrator declarations with multiple baseline items within R_{crit}, P_{group} credits the demonstrator with detecting a single group of targets, and does not penalize for any missed detections. P_{near} measures the probability that an item will be found if holes of radius R_{crit} are dug on all demonstrator declarations, allowing multiple baseline targets to be matched to a single demonstrator declaration. The following table compares detection capability calculated using these three measures, where P_{group} and P_{match} represent upper and lower bounds on the probability of detection, P_d, achieved on this demonstration. The false alarm rate, *FAR*, will be the same for all three measures; the false negative ratio will be different. Also presented in the table are P_{match} and P_{near} calculated with the plastic mines removed from the baseline. All detection measures are for ordnance items only.

Calculation method	Value
P_{match}[1]	0.04
P_{match}(no mines)	0.04
P_{group}	0.04
P_{near}	0.04
P_{near} (no mines)	0.04

[1] For this demonstration, the correction to detection capabilities for random matches was small (< 0.03).

The following tables provide information about specific demonstrator declarations and baseline items which may be of use to the demonstrator in assessing target identification capabilities for future operations, or for making improvements in the sensor or the data processing algorithms. The correct locations and types are provided for 17 baseline items that were dug up during or shortly after the remediation demonstrations. Five of these baseline items are isolated, individual objects. The remaining 12 are six closely spaced pairs of items. These items have either been removed from the site or were emplaced again at new locations. Three demonstrator declarations that were not within 2 meters of a baseline item are also provided.

Locations and Types of Selected Baseline Targets				
Type	Size	X[a]	Y	Depth
Bomb	Large	2104200.6	14139531.2	8.0
Bomb	Large	2105272.6	14139550.2	5.6
Bomb	Large	2105272.5	14139555.1	2.9
Projectile	Medium	2105087.4	14139006.9	3.4
Projectile	Medium	2105105.8	14139336.8	2.7
Projectile	Small	2105134.5	14139561.0	6.1
Projectile	Medium	2105326.1	14139555.1	2.0
Projectile	Medium	2105014.3	14139251.5	3.3
Mortar	Small	2105139.5	14139036.9	3.1
Mortar	Small	2105327.9	14139554.6	1.2
Mortar	Small	2104978.4	14139240.2	1.3
Mortar	Small	2104976.0	14139229.1	0.8
Mortar	Small	2104976.7	14139229.3	0.9
Mortar	Small	2104013.6	14138535.2	1.1
Mortar	Small	2104017.5	14138533.9	0.9
NonOrd	Large	2105014.7	14139251.1	0.9
NonOrd	Large	2105139.8	14139037.1	1.1

[a] All dimensions are in feet.

Demonstrator Declarations Not Within 2 m of a Baseline Item	
1[a]	No known baseline item within 2 m of this location.
10	No known baseline item within 2 m of this location.
218	No known baseline item within 2 m of this location.
[a] Target numbers are from the demonstrator numbering scheme.	

III. SUMMARY

The purpose of the 1994 Jefferson Proving Ground demonstration was to evaluate the state of the art in UXO detection and remediation technology. For all detection demonstrators, the maximum detection capabilities as measured by the target matching algorithm have been lower than P_{match} = 0.65 and there have been multiple false declarations per ordnance item detected. Further, statistical analysis of the JPG data indicates that demonstrators with similar detection capabilities cannot be ordinally ranked with confidence. Therefore, demonstrators are binned on the basis of P_{match}, false alarm rate, and a separation criteria d, which combines measures of P_{match} and P_{fa} into a single variable and allows for comparison of demonstrators with widely different detection capabilities and false alarm rates. In this assessment, using the R_{crit} = 2.0 m numbers, ENSCO performed significantly worse than the average JPG demonstrator in terms of P_{match}, in the third of four bins in false alarms, and in the last of four bins in the separation measure.

This page is intentionally left blank

EVALUATION OF EODT SERVICES
40-ACRE SITE

I. DEMONSTRATOR RESULTS

	Entire Site	Area Searched[1]
Detection Capability		
overall detection ratio, R_{all}	0.04	0.07
ordnance detection ratio, R_{ord}	0.05	0.07
nonordnance detection ratio, R_{nonord}	0.03	0.06
mistyped ordnance ratio, *MR*	0.0‡	0.0‡
False Positive Ratio, *FPR*	1.0‡	1.0‡
False Negative Ratio, *FNR*	0.80	0.80
False Alarm Rate, *FAR*	0.19	0.42
Radial Accuracy, ΔR_{xy} *(feet)*	4.2	4.2
Depth Accuracy, ΔR_z *(feet)*	**	**
Target Classification Capabilities		
single targets	1.0	1.0
multiple targets	*	*
bombs	0	0
projectiles	0	0
mortars	0	0
mines	*	*
clusters	*	*

1 The portion of the site searched by this demonstrator may not be representative of the overall difficulty of the entire 40-acre site.

* Ratio is undefined (i.e., zero denominator). A classification ratio is undefined if no targets of that class are located by the demonstrator.

‡ The values shown here are a consequence of the demonstrator declaring all items as ordnance.

** No depth information provided.

Survey Rate

The demonstrator searched approximately 11 acres of the 40-acre site within the allotted 1 week. The proposal gave survey rates as 3 acres per day for open, flat terrain.

II. DISCUSSION OF DEMONSTRATOR PERFORMANCE

EODT used a Schonstedt magnetometer and an EM-31 conductivity sensor for detection, and a GEODAPS differential GPS for navigation. The Schonstedt magnetometer has a range of 2–5 feet, and the conductivity sensor has a range of 10–15 feet. The purpose of this demonstration was to assess the capabilities of the DANS (Data Acquisition and Navigation System).

This demonstrator did not attempt to determine whether items were ordnance or nonordnance, typing all detections as ordnance items. Because no attempt was made to distinguish ordnance from nonordnance, several of the ratios used in this evaluation are not useful for characterizing this demonstrator. For example, $MR = 0$ not because the demonstrator made no errors in separating ordnance from nonordnance; rather, because no items were typed as nonordnance, no ordnance items could be misidentified as nonordnance. $FPR = 1.0$ since any nonordnance items found were typed as ordnance.

The demonstrator surveyed approximately 11 acres of the 40-acre site. When the demonstrator's detection capability and false alarm rates are recalculated to include only the portion of the site searched, the results are as follows:

$$P_{match} = 0.07$$

$$P_{match}(\text{ord}) = 0.07$$

$$FAR = 0.42$$

Corrections for random hits by this demonstrator reduced the uncorrected P_{match} = 0.07 to a corrected P_{match} = 0.06, where P_{match} refers to the ordnance detection capability on only the area searched.

Detection capabilities were calculated for the various ordnance types emplaced. Only the targets in the area searched were considered, and credit was given for a detection regardless of classification ability, when calculating the detection capabilities reported in this paragraph. This demonstrator's best performance was in the detection of projectiles and mortars, with P_{match} =0.14 and 0.11. This demonstrator's worst performance was in detecting bombs, mines, and cluster targets, with zero detections of each target type. Since

the sensors used by this demonstrator can detect only metal objects, the inability to detect plastic mines is to be expected. The performance of this demonstrator was best at intermediate depths. When the targets are grouped as those buried less than 3 feet, between 3 and 6 feet, and more than 6 feet deep, P_{match} was 0.05, 0.15, and 0.0, respectively. The demonstrator was best able to detect medium targets, with P_{match} = 0.13, and detected large and small targets at P_{match} = 0.05 and P_{match} = 0.06, respectively.

The detection capabilities calculated above are all based on P_{match}, the probability of a one-to-one match between a demonstrator declaration and an emplaced target. This method of determining detection probability assumes a high degree of sensor resolution, which may not be justified if a sensor cannot resolve signals resulting from two (or more) closely spaced objects. Two alternative measures of detection capability are also considered. For demonstrator declarations with multiple baseline items within R_{crit}, P_{group} credits the demonstrator with detecting a single group of targets, and does not penalize for any missed detections. P_{near} measures the probability that an item will be found if holes of radius R_{crit} are dug on all demonstrator declarations, allowing multiple baseline targets to be matched to a single demonstrator declaration. The following table compares detection capability calculated using these three measures, where P_{group} and P_{match} represent upper and lower bounds on the probability of detection, P_d, achieved on this demonstration. The false alarm rate, *FAR*, will be the same for all three measures; the false negative ratio will be different. Also presented in the table are P_{match} and P_{near} calculated with the plastic mines removed from the baseline. All detection measures are for ordnance items only.

Calculation method	Value
P_{match}[1]	0.07
P_{match}(no mines)	0.10
P_{group}	0.07
P_{near}	0.10
P_{near} (no mines)	0.14

[1] For this demonstration, the correction to detection capabilities for random matches was small (< 0.03).

The following tables provide information about specific demonstrator declarations and baseline items which may be of use to the demonstrator in assessing target identification capabilities for future operations, or for making improvements in the sensor or the data processing algorithms. The correct locations and types are provided for 17 baseline items that were dug up during or shortly after the remediation demonstrations. Five of

these baseline items are isolated, individual objects. The remaining 12 are six closely spaced pairs of items. These items have either been removed from the site or were emplaced again at new locations. Three demonstrator declarations that were not within 2 meters of a baseline item are also provided.

Locations and Types of Selected Baseline Targets				
Type	Size	X[a]	Y	Depth
Bomb	Large	2104200.6	14139531.2	8.0
Bomb	Large	2105272.6	14139550.2	5.6
Bomb	Large	2105272.5	14139555.1	2.9
Projectile	Medium	2105087.4	14139006.9	3.4
Projectile	Medium	2105105.8	14139336.8	2.7
Projectile	Small	2105134.5	14139561.0	6.1
Projectile	Medium	2105326.1	14139555.1	2.0
Projectile	Medium	2105014.3	14139251.5	3.3
Mortar	Small	2105139.5	14139036.9	3.1
Mortar	Small	2105327.9	14139554.6	1.2
Mortar	Small	2104978.4	14139240.2	1.3
Mortar	Small	2104976.0	14139229.1	0.8
Mortar	Small	2104976.7	14139229.3	0.9
Mortar	Small	2104013.6	14138535.2	1.1
Mortar	Small	2104017.5	14138533.9	0.9
NonOrd	Large	2105014.7	14139251.1	0.9
NonOrd	Large	2105139.8	14139037.1	1.1
[a] All dimensions are in feet.				

Demonstrator Declarations Not Within 2 m of a Baseline Item	
5[a]	No known baseline item within 2 m of this location.
37	No known baseline item within 2 m of this location.
42	No known baseline item within 2 m of this location.
[a] Target numbers are from the demonstrator numbering scheme.	

III. SUMMARY

The purpose of the 1994 Jefferson Proving Ground demonstration was to evaluate the state of the art in UXO detection and remediation technology. For all detection demonstrators, the maximum detection capabilities as measured by the target matching algorithm have been lower than $P_{match} = 0.65$ and there have been multiple false declarations per ordnance item detected. Further, statistical analysis of the JPG data indicates that demonstrators with similar detection capabilities cannot be ordinally ranked with confidence. Therefore, demonstrators are binned on the basis of P_{match}, false alarm rate, and a separation criteria d, which combines measures of P_{match} and P_{fa} into a single variable and allows for comparison of demonstrators with widely different detection capabilities and false alarm rates. In this assessment, EODT performed significantly worse than the average JPG demonstrator in terms of P_{match}, in the top bin of four bins in false alarms, and in the third of four bins in the separation measure.

This page is intentionally left blank

EVALUATION OF FOERSTER INSTRUMENTS
40-ACRE SITE

I. DEMONSTRATOR RESULTS

	Entire Site	Area Searched[1]
Detection Capability		
overall detection ratio, R_{all}	0.27	0.40
ordnance detection ratio, R_{ord}	0.26	0.38
nonordnance detection ratio, R_{nonord}	0.30	0.44
mistyped ordnance ratio, *MR*	0.0	0.0
False Positive Ratio, *FPR*	1.0	1.0
False Negative Ratio, *FNR*	0.87	0.87
False Alarm Rate, *FAR*	2.0	3.23
Radial Accuracy, ΔR_{xy} *(feet)*	3.1	3.0
Depth Accuracy, ΔR_z *(feet)*	2.5	2.0
Target Classification Capabilities		
single targets	0.96	0.95
multiple targets	*	*
bombs	0	0
projectiles	0.37	0.37
mortars	0.18	0.18
mines	*	*
clusters	*	*

[1] The portion of the site searched by this demonstrator may not be representative of the overall difficulty of the entire 40-acre site.

* Ratio is undefined (i.e., zero denominator). A classification ratio is undefined if no targets of that class are located by the demonstrator.

Survey Rate

The demonstrator searched approximately 24 acres of the 40-acre site within the allotted 1 week. The demonstrator estimated in the proposal that 2.5 to 3 days would be required to survey the site. The demonstrator report does not indicate the reason that the

survey rate was much lower than expected, commenting only that the remainder of the site could not be searched "due to time limitations."

II. DISCUSSION OF DEMONSTRATOR PERFORMANCE

Foerster Instruments used a Ferex (Mark 26) Standard Sensor, a Ferex Deep Search Sensor, and a Minex 2FD Standard Sensor. The Ferex sensors are ferrous detecting gradient magnetometers, and the Minex is an induction coil sensor that can detect any metal object. The sensors were mounted on a surface towed platform. Terrain at JPG forced the Minex platform to operate at a stand-off height approximately 12 inches from the ground, reducing the sensitivity that would be available operating at the normal standoff height of 2–3 inches. Differential GPS was used for navigation.

The demonstrator surveyed approximately 21 acres of the 40-acre site. When the demonstrator's detection capability and false alarm rates are recalculated to include only the portion of the site searched, the results are as follows:

$$P_{match} = 0.40$$

$$P_{match}\,(\text{ord}) = 0.38$$

$$FAR = 3.2$$

Corrections for random hits by this demonstrator reduced the uncorrected P_{match} = 0.38 to a corrected P_{match} = 0.35, where P_{match} refers to the ordnance detection capability on only the area searched.

Detection capabilities were calculated for the various ordnance items emplaced. Only the targets in the area searched were considered and credit was given for a detection regardless of classification ability, when calculating the detection capabilities reported in this paragraph. This demonstrator's best performance was in the detection of bombs, with P_{match} = 0.76. This demonstrator's worst performance was in detecting mines and cluster targets, with zero detections of either target type. Since the sensors used by this demonstrator can detect only metal objects, the inability to detect plastic mines is to be expected. The performance of this demonstrator was best at deep and intermediate depths. When the targets are grouped as those buried less than 3 feet, between 3 and 6 feet, and more than 6 feet deep, P_{match} was 0.30, 0.56, and 0.62, respectively. The demonstrator was best able to detect medium size targets, with P_{match} = 0.63, and detected small and large targets at P_{match} = 0.25 and P_{match} = 0.50, respectively.

The detection capabilities calculated above are all based on P_{match}, the probability of a one-to-one match between a demonstrator declaration and an emplaced target. This method of determining detection probability assumes a high degree of sensor resolution, which may not be justified if a sensor cannot resolve signals resulting from two (or more) closely spaced objects. Two alternative measures of detection capability are also considered. For demonstrator declarations with multiple baseline items within R_{crit}, P_{group} credits the demonstrator with detecting a single group of targets, and does not penalize for any missed detections. P_{near} measures the probability that an item will be found if holes of radius R_{crit} are dug on all demonstrator declarations, allowing multiple baseline targets to be matched to a single demonstrator declaration. The following table compares detection capability calculated using these three measures, where P_{group} and P_{match} represent upper and lower bounds on the probability of detection, P_d, achieved on this demonstration. The false alarm rate, *FAR*, will be the same for all three measures; the false negative ratio will be different. Also presented in the table are P_{match} and P_{near} calculated with the plastic mines removed from the baseline. All detection measures are for ordnance items only.

Calculation method	Value
P_{match}[1]	0.38
P_{match}(no mines)	0.48
P_{group}	0.43
P_{near}	0.49
P_{near} (no mines)	0.61

[1] For this demonstration, the correction to detection capabilities for random matches was small (< 0.03).

The following tables provide information about specific demonstrator declarations and baseline items which may be of use to the demonstrator in assessing target identification capabilities for future operations, or for making improvements in the sensor or the data processing algorithms. The correct locations and types are provided for 17 baseline items that were dug up during or shortly after the remediation demonstrations. Five of these baseline items are isolated, individual objects. The remaining 12 are six closely spaced pairs of items. These items have either been removed from the site or were emplaced again at new locations. Three demonstrator declarations that were not within 2 meters of a baseline item are also provided.

Locations and Types of Selected Baseline Targets				
Type	Size	X[a]	Y	Depth
Bomb	Large	2104200.6	14139531.2	8.0
Bomb	Large	2105272.6	14139550.2	5.6
Bomb	Large	2105272.5	14139555.1	2.9
Projectile	Medium	2105087.4	14139006.9	3.4
Projectile	Medium	2105105.8	14139336.8	2.7
Projectile	Small	2105134.5	14139561.0	6.1
Projectile	Medium	2105326.1	14139555.1	2.0
Projectile	Medium	2105014.3	14139251.5	3.3
Mortar	Small	2105139.5	14139036.9	3.1
Mortar	Small	2105327.9	14139554.6	1.2
Mortar	Small	2104978.4	14139240.2	1.3
Mortar	Small	2104976.0	14139229.1	0.8
Mortar	Small	2104976.7	14139229.3	0.9
Mortar	Small	2104013.6	14138535.2	1.1
Mortar	Small	2104017.5	14138533.9	0.9
NonOrd	Large	2105014.7	14139251.1	0.9
NonOrd	Large	2105139.8	14139037.1	1.1

[a] All dimensions are in feet.

Demonstrator Declarations Not Within 2 m of a Baseline Item	
5[a]	No known baseline item within 2 m of this location.
12	No known baseline item within 2 m of this location.
409	No known baseline item within 2 m of this location.

[a] Target numbers are from the demonstrator numbering scheme.

III. SUMMARY

The purpose of the 1994 Jefferson Proving Ground demonstration was to evaluate the state of the art in UXO detection and remediation technology. For all detection demonstrators, the maximum detection capabilities as measured by the target matching algorithm have been lower than P_{match} = 0.65 and there have been multiple false declarations per ordnance item detected. Further, statistical analysis of the JPG data

indicates that demonstrators with similar detection capabilities cannot be ordinally ranked with confidence. Therefore, demonstrators are binned on the basis of P_{match}, false alarm rate, and a separation criteria d, which combines measures of P_{match} and P_{fa} into a single variable and allows for comparison of demonstrators with widely different detection capabilities and false alarm rates. In this assessment, Foerster performed significantly better than the average JPG demonstrator in terms of P_{match}, in the third of four bins in false alarms, and in the third of four bins in the separation measure.

This page is intentionally left blank

EVALUATION OF GDE SYSTEMS
40-ACRE SITE

I. DEMONSTRATOR RESULTS

	Entire Site	Area Searched[1]
Detection Capability		
overall detection ratio, R_{all}	0.05	0.40
ordnance detection ratio, R_{ord}	0.05	0.39
nonordnance detection ratio, R_{nonord}	0.04	0.43
mistyped ordnance ratio, *MR*	0.0‡	0.0‡
False Positive Ratio, *FPR*	1.0‡	1.0‡
False Negative Ratio, *FNR*	0.99	0.99
False Alarm Rate, *FAR*	3.93	29.7
Radial Accuracy, ΔR_{xy} *(feet)*	4.6	4.6
Depth Accuracy, ΔR_z *(feet)*	6.8	6.8
Target Classification Capabilities		
single targets	1.0	1.0
multiple targets	0.0	0.0
bombs	1.0	1.0
projectiles	0.0	0.0
mortars	0.0	0.0
mines	*	*
clusters	0	0.0

[1] The portion of the site searched by this demonstrator may not be representative of the overall difficulty of the entire 40-acre site. The area of the site that this demonstrator reported visiting was not consistent with the location of his target declarations. Therefore, the area searched values are calculated considering only the grid cells that contained demonstrator declarations.

* Ratio is undefined (i.e., zero denominator). A classification ratio is undefined if no targets of that class are located by the demonstrator.

‡ The values shown here are a consequence of the demonstrator declaring all items as ordnance.

Survey Rate

The demonstrator searched approximately 6.9 acres of the 40-acre site within the allotted 1 week. The proposal gave the survey rate as 2.2 acres per hour, with expected completion of the site in 4 days.

II. DISCUSSION OF DEMONSTRATOR RESULTS

GDE Systems used a prototype surface-towed imaging ground penetrating radar sensor with a range of 10 to 15 feet for detection of UXO and a rope/tape/odometer navigation system. The demonstrator report noted that the disking of the soil left large diameter clods of earth, and that there were standing puddles of water and mud holes. The proposal lists both soil moisture and surface roughness as conditions that are expected to stress the GDE system. The target list submitted by this demonstrator contained target identification numbers that were used more than once, making the input data incompatible with the target matching algorithm. Therefore, the target identification numbers of this demonstrator were changed to eliminate duplications prior to analysis.

This demonstrator did not attempt to determine whether items were ordnance or nonordnance, typing all detections as ordnance items. Because no attempt was made to distinguish ordnance from nonordnance, several of the ratios used in this evaluation are not useful for characterizing this demonstrator. For example, $MR = 0$ not because the demonstrator made no errors in separating ordnance from nonordnance; rather, because no items were typed as non-ordnance, no ordnance items could be misidentified as nonordnance. $FPR = 1.0$ since any nonordnance items found were typed as ordnance. Of particular note for this demonstrator is the false negative ratio of 0.99. As a practical matter, if all the ordnance declarations of this demonstrator were investigated, some of the holes dug would contain emplaced nonordnance items, and 99% of the remaining holes dug for remediation would not contain any of the ordnance items known by the government to be present.

The demonstrator surveyed approximately 6.9 acres of the 40-acre site. When the demonstrator's detection capability and false alarm rates are recalculated to include only the portion of the site searched, the results are as follows:

$$P_{match} = 0.39$$

$$FAR = 29.7$$

The false alarm rate on the area searched by this demonstrator is by far the highest experienced by any demonstrator and is approximately an order of magnitude greater than the average false alarm rate for the demonstrators on the 40-acre site.

Corrections for random hits by this demonstrator reduced the uncorrected P_{match} = 0.39 to a corrected P_{match} = – 0.02, where P_{match} refers to the ordnance detection capability on only the area searched.

Detection capabilities were calculated for the various ordnance types emplaced. Only the targets in the area searched were considered and credit was given for a detection regardless of classification ability, when calculating the detection capabilities reported in this paragraph. This demonstrator's best performance was in the detection of bombs, with P_{match} = 0.33. This demonstrator's worst performance was in detecting mines, with zero detections in this category. The performance of this demonstrator was best for deep targets. When the targets are grouped as those buried less than 3 feet, between 3 and 6 feet, and more than 6 feet deep, P_{match} was 0.38, 0.29, and 0.57, respectively. The demonstrator was best able to detect medium targets, with P_{match} = 0.50, and detected small and large targets at P_{match} = 0.37 and P_{match} = 0.40, respectively.

The detection capabilities calculated above are all based on P_{match}, the probability of a one-to-one match between a demonstrator declaration and an emplaced target. This method of determining detection probability assumes a high degree of sensor resolution, which may not be justified if a sensor cannot resolve signals resulting from two (or more) closely spaced objects. Two alternative measures of detection capability are also considered. For demonstrator declarations with multiple baseline items within R_{crit}, P_{group} credits the demonstrator with detecting a single group of targets, and does not penalize for any missed detections. P_{near} measures the probability that an item will be found if holes of radius R_{crit} are dug on all demonstrator declarations, allowing multiple baseline targets to be matched to a single demonstrator declaration. The following table compares detection capability calculated using these three measures, where P_{group} and P_{match} represent upper and lower bounds on the probability of detection, P_d, achieved on this demonstration. The false alarm rate, *FAR*, will be the same for all three measures; the false negative ratio will be different. Also presented in the table are P_{match} and P_{near} calculated with the plastic mines removed from the baseline. All detection measures are for ordnance items only.

Calculation method	Value
P_{match}[1]	0.39
P_{match}(no mines)	0.39
P_{group}	0.39
P_{near}	0.39
P_{near} (no mines)	0.39

[1] For this demonstration, the correction to detection capabilities for random matches reduced the P_{match} reported here to –0.02.

The following tables provide information about specific demonstrator declarations and baseline items which may be of use to the demonstrator in assessing target identification capabilities for future operations, or for making improvements in the sensor or the data processing algorithms. The correct locations and types are provided for 17 baseline items that were dug up during or shortly after the remediation demonstrations. Five of these baseline items are isolated, individual objects. The remaining 12 are six closely spaced pairs of items. These items have either been removed from the site or were emplaced again at new locations. Three demonstrator declarations that were not within 2 meters of a baseline item are also provided.

Locations and Types of Selected Baseline Targets				
Type	Size	X[a]	Y	Depth
Bomb	Large	2104200.6	14139531.2	8.0
Bomb	Large	2105272.6	14139550.2	5.6
Bomb	Large	2105272.5	14139555.1	2.9
Projectile	Medium	2105087.4	14139006.9	3.4
Projectile	Medium	2105105.8	14139336.8	2.7
Projectile	Small	2105134.5	14139561.0	6.1
Projectile	Medium	2105326.1	14139555.1	2.0
Projectile	Medium	2105014.3	14139251.5	3.3
Mortar	Small	2105139.5	14139036.9	3.1
Mortar	Small	2105327.9	14139554.6	1.2
Mortar	Small	2104978.4	14139240.2	1.3
Mortar	Small	2104976.0	14139229.1	0.8
Mortar	Small	2104976.7	14139229.3	0.9
Mortar	Small	2104013.6	14138535.2	1.1
Mortar	Small	2104017.5	14138533.9	0.9
NonOrd	Large	2105014.7	14139251.1	0.9
NonOrd	Large	2105139.8	14139037.1	1.1
[a] All dimensions are in feet.				

Demonstrator Declarations Not Within 2 m of a Baseline Item	
5[a]	No known baseline item within 2 m of this location.
13	No known baseline item within 2 m of this location.
102	No known baseline item within 2 m of this location.
[a] Target numbers are from the demonstrator numbering scheme.	

III. SUMMARY

The purpose of the 1994 Jefferson Proving Ground demonstration was to evaluate the state of the art in UXO detection and remediation technology. For all detection demonstrators, the maximum detection capabilities as measured by the target matching algorithm have been lower than $P_{match} = 0.65$ and there have been multiple false declarations per ordnance item detected. Further, statistical analysis of the JPG data indicates that demon-

strators with similar detection capabilities cannot be ordinally ranked with confidence. Therefore, demonstrators are binned on the basis of P_{match}, false alarm rate, and a separation criteria d, which combines measures of P_{match} and P_{fa} into a single variable and allows for comparison of demonstrators with widely different detection capabilities and false alarm rates. In this assessment, GDE performed significantly better than the average JPG demonstrator in terms of P_{match}, in the last of four bins in false alarms, and in the last of four bins in the separation measure.

EVALUATION OF GEO-CENTERS, INC.
40-ACRE SITE

I. DEMONSTRATOR RESULTS

	Entire Site	Area Searched[1]
Detection Capability		
overall detection ratio, R_{all}	0.47	0.47
ordnance detection ratio, R_{ord}	0.46	0.46
nonordnance detection ratio, R_{nonord}	0.50	0.50
mistyped ordnance ratio, *MR*	0.0	0.0
False Positive Ratio, *FPR*	1.0	1.0
False Negative Ratio, *FNR*	0.75	0.75
False Alarm Rate, *FAR*	1.33	1.33
Radial Accuracy, ΔR_{xy} *(feet)*	3.0	3.0
Depth Accuracy, ΔR_z *(feet)*	2.1	2.1
Target Classification Capabilities		
single targets	1.0	1.0
multiple targets	0.0	0.0
bombs	0.82	0.82
projectiles	0.83	0.83
mortars	0.0	0.0
mines	*	*
clusters	0.0	0.0

[1] These columns are identical because this demonstrator searched the entire 40-acre site.

* Ratio is undefined (i.e., zero denominator). A classification ratio is undefined if no targets of that class are detected by the demonstrator.

Survey Rate

The demonstrator searched the entire 40-acre site within the allotted 1 week. The proposal gave the survey rate as 20 acres per day for the surface towed magnetometer.

II. DISCUSSION OF DEMONSTRATOR PERFORMANCE

Geo-Centers used a Surface Towed Ordnance Locator System (STOLS) for the JPG demonstration. This system consisted of two Gmtrcs/Scntx magnetometers and one Foerster hybrid magnetometer/gradiometer detector. The Gmtrcs/Scntx had a range of more than 25 feet, and the Foerster had a range of 5–10 feet. Areas of the site inaccessible to the STOLS were surveyed using a hand carried system. Differential GPS was used for navigation. Per the proposal, the demonstrator expected a detection efficiency of > 90 % with one-half meter accuracy, and a survey rate of approximately 20 acres per day.

Geo-Centers searched the entire 40-acre site. Corrections for random hits by this demonstrator reduced the uncorrected $P_{match} = 0.46$ to a corrected $P_{match} = 0.45$, where P_{match} refers to the ordnance detection capability.

Detection capabilities were calculated for the various ordnance types emplaced. Credit was given for a detection regardless of classification ability when calculating the detection capabilities reported in this paragraph. This demonstrator's best performance was in the detection of bombs, with $P_{match} = 0.84$. This demonstrator's worst performance was in detecting mines, with zero detections of this target type. Since the sensors used by this demonstrator can detect only metal objects, the inability to detect plastic mines is to be expected. The performance of this demonstrator was best in detection of deep targets. When the targets are grouped as those buried less than 3 feet, between 3 and 6 feet, and more than 6 feet deep, P_{match} was 0.39, 0.60, and 0.72, respectively. The demonstrator was best able to detect large and medium size targets, with $P_{match} = 0.62$ and $P_{match} = 0.65$, respectively, and detected small targets at $P_{match} = 0.30$.

The detection capabilities calculated above are all based on P_{match}, the probability of a one-to-one match between a demonstrator declaration and an emplaced target. This method of determining detection probability assumes a high degree of sensor resolution, which may not be justified if a sensor cannot resolve signals resulting from two (or more) closely spaced objects. Two alternative measures of detection capability are also considered. For demonstrator declarations with multiple baseline items within R_{crit}, P_{group} credits the demonstrator with detecting a single group of targets, and does not penalize for any missed detections. P_{near} measures the probability that an item will be found if holes of radius R_{crit} are dug on all demonstrator declarations, allowing multiple baseline targets to be matched to a single demonstrator declaration. The following table compares detection capability calculated using these three measures, where P_{group} and P_{match} represent upper and lower bounds on the probability of detection, P_d, achieved on this demonstration. The

false alarm rate, *FAR*, will be the same for all three measures; the false negative ratio will be different. Also presented in the table are P_{match} and P_{near} calculated with the plastic mines removed from the baseline. All detection measures are for ordnance items only.

Calculation method	Value
P_{match}[1]	0.46
P_{match}(no mines)	0.53
P_{group}	0.53
P_{near}	0.60
P_{near} (no mines)	0.69

[1] For this demonstration, the correction to detection capabilities for random matches was small (< 0.03).

The following tables provide information about specific demonstrator declarations and baseline items which may be of use to the demonstrator in assessing target identification capabilities for future operations, or for making improvements in the sensor or the data processing algorithms. The correct locations and types are provided for 17 baseline items that were dug up during or shortly after the remediation demonstrations. Five of these baseline items are isolated, individual objects. The remaining 12 are six closely spaced pairs of items. These items have either been removed from the site or were emplaced again at new locations. Three demonstrator declarations that were not within 2 meters of a baseline item are also provided.

Locations and Types of Selected Baseline Targets				
Type	Size	X[a]	Y	Depth
Bomb	Large	2104200.6	14139531.2	8.0
Bomb	Large	2105272.6	14139550.2	5.6
Bomb	Large	2105272.5	14139555.1	2.9
Projectile	Medium	2105087.4	14139006.9	3.4
Projectile	Medium	2105105.8	14139336.8	2.7
Projectile	Small	2105134.5	14139561.0	6.1
Projectile	Medium	2105326.1	14139555.1	2.0
Projectile	Medium	2105014.3	14139251.5	3.3
Mortar	Small	2105139.5	14139036.9	3.1
Mortar	Small	2105327.9	14139554.6	1.2
Mortar	Small	2104978.4	14139240.2	1.3
Mortar	Small	2104976.0	14139229.1	0.8
Mortar	Small	2104976.7	14139229.3	0.9
Mortar	Small	2104013.6	14138535.2	1.1
Mortar	Small	2104017.5	14138533.9	0.9
NonOrd	Large	2105014.7	14139251.1	0.9
NonOrd	Large	2105139.8	14139037.1	1.1
[a] All dimensions are in feet.				

Demonstrator Declarations Not Within 2 m of a Baseline Item	
4[a]	No known baseline item within 2 m of this location.
10	No known baseline item within 2 m of this location.
31	No known baseline item within 2 m of this location.
[a] Target numbers are from the demonstrator numbering scheme.	

III. SUMMARY

The purpose of the 1994 Jefferson Proving Ground demonstration was to evaluate the state of the art in UXO detection and remediation technology. For all detection demonstrators, the maximum detection capabilities as measured by the target matching algorithm have been lower than $P_{match} = 0.65$, and there have been multiple false declarations per ordnance item detected. Further, statistical analysis of the JPG data indicates that

demonstrators with similar detection capabilities cannot be ordinally ranked with confidence. Therefore, demonstrators are binned on the basis of P_{match}, false alarm rate, and a separation criteria d, which combines measures of P_{match} and P_{fa} into a single variable and allows for comparison of demonstrators with widely different detection capabilities and false alarm rates. In this assessment, Geo-Centers performed significantly better than the average JPG demonstrator in terms of P_{match}, in the second of four bins in false alarms, and in the top bin of four bins in the separation measure.

This page is intentionally left blank

EVALUATION OF GEOMETRICS, INC.
40-ACRE SITE

I. DEMONSTRATOR RESULTS

	Entire Site	Area Searched[1]
Detection Capability		
overall detection ratio, R_{all}	0.21	0.23
ordnance detection ratio, R_{ord}	0.22	0.23
nonordnance detection ratio, R_{nonord}	0.18	0.23
mistyped ordnance ratio, *MR*	0.0	0.0
False Positive Ratio, *FPR*	1.0	1.0
False Negative Ratio, *FNR*	0.74	0.74
False Alarm Rate, *FAR*	0.63	0.70
Radial Accuracy, ΔR_{xy} *(feet)*	4.34	4.32
Depth Accuracy, ΔR_z *(feet)*	2.6	2.57
Target Classification Capabilities		
single targets	0.75	0.75
multiple targets	1.0	1.0
bombs	0.45	0.45
projectiles	0.95	0.95
mortars	0	0
mines	*	*
clusters	0	0

[1] The portion of the site searched by this demonstrator may not be representative of the overall difficulty of the entire 40-acre site.

* Ratio is undefined (i.e., zero denominator). A classification ratio is undefined if no targets of that class are located by the demonstrator.

Survey Rate

The demonstrator searched approximately 35 acres of the 40-acre site within the allotted 1 week. The proposal gave survey rates as 10 acres per day.

II. DISCUSSION OF DEMONSTRATOR PERFORMANCE

Geometrics used a prototype Geometrics MagDIS man-portable system consisting of five cesium-vapor magnetometer sensors for detection. Differential GPS was used for navigation. The demonstrator report comments on the terrain difficulties caused by the large clumps of earth, and the cut-off and turned up roots caused by disking the site.

Corrections for random hits by this demonstrator reduced the uncorrected P_{match} = 0.23 to a corrected P_{match} = 0.22, where P_{match} refers to the ordnance detection capability on only the area searched.

A number of the items that the demonstrator classified as ordnance items were indicated in the comment fields as possible trench, pipe, anomaly, or construction equipment. The P_{match}, *FNR*, *FPR*, and *FAR* for this demonstrator were recalculated after taking out all such declarations. The corrected values are:

$$P_{match} = 0.20$$

$$P_{match}(\text{ord}) = 0.21$$

$$FNR = 0.66$$

$$FPR = 1.0$$

$$FAR = 0.43$$

where P_{match} is the overall detection ratio on only the area searched. While the demonstrator's P_{match} did not go down significantly, the *FNR* and *FAR* are greatly reduced. The measures used in the remainder of this report are calculated with these declarations removed from the demonstrator data set.

Detection capability was calculated for the various ordnance types emplaced. Only the targets in the area searched were considered, and credit was given for a detection regardless of classification ability, when calculating the detection capabilities reported in this paragraph. This demonstrator's best performance was in the detection of bombs, with P_{match} = 0.53. This demonstrator's worst performance was in the detection of mines, with zero detections, and mortars with P_{match} = 0.10. Since the sensors used by this demonstrator can detect only metal objects, the inability to detect plastic mines is to be expected. The performance of this demonstrator was best for deep targets. When the targets are grouped as those buried less than 3 feet, between 3 and 6 feet, and more than 6 feet deep, P_{match} was 0.13, 0.17, and 0.63, respectively. The demonstrator was best able to detect

large targets, with $P_{match} = 0.35$, and detected small and medium targets at $P_{match} = 0.08$ and $P_{match} = 0.26$, respectively.

The detection capabilities calculated above are all based on P_{match}, the probability of a one-to-one match between a demonstrator declaration and an emplaced target. This method of determining detection probability assumes a high degree of sensor resolution, which may not be justified if a sensor cannot resolve signals resulting from two (or more) closely spaced objects. Two alternative measures of detection capability are also considered. For demonstrator declarations with multiple baseline items within R_{crit}, P_{group} credits the demonstrator with detecting a single group of targets, and does not penalize for any missed detections. P_{near} measures the probability that an item will be found if holes of radius R_{crit} are dug on all demonstrator declarations, allowing multiple baseline targets to be matched to a single demonstrator declaration. The following table compares detection capability calculated using these three measures, where P_{group} and P_{match} represent upper and lower bounds on the probability of detection, P_d, achieved on this demonstration. The false alarm rate, *FAR*, will be the same for all three measures; the false negative ratio will be different. Also presented in the table are P_{match} and P_{near} calculated with the plastic mines removed from the baseline. All detection measures are for ordnance items only.

Calculation method	Value
P_{match}[1]	0.23
P_{match}*(no mines)*	0.24
P_{group}	0.21
P_{near}	0.23
P_{near} *(no mines)*	0.27

[1] For this demonstration, the correction to detection capabilities for random matches was small (< 0.03).

The following tables provide information about specific demonstrator declarations and baseline items which may be of use to the demonstrator in assessing target identification capabilities for future operations, or for making improvements in the sensor or the data processing algorithms. The correct locations and types are provided for 17 baseline items that were dug up during or shortly after the remediation demonstrations. Five of these baseline items are isolated, individual objects. The remaining 12 are six closely spaced pairs of items. These items have either been removed from the site or were emplaced again at new locations. Three demonstrator declarations that were not within 2 meters of a baseline item are also provided.

Locations and Types of Selected Baseline Targets				
Type	Size	X[a]	Y	Depth
Bomb	Large	2104200.6	14139531.2	8.0
Bomb	Large	2105272.6	14139550.2	5.6
Bomb	Large	2105272.5	14139555.1	2.9
Projectile	Medium	2105087.4	14139006.9	3.4
Projectile	Medium	2105105.8	14139336.8	2.7
Projectile	Small	2105134.5	14139561.0	6.1
Projectile	Medium	2105326.1	14139555.1	2.0
Projectile	Medium	2105014.3	14139251.5	3.3
Mortar	Small	2105139.5	14139036.9	3.1
Mortar	Small	2105327.9	14139554.6	1.2
Mortar	Small	2104978.4	14139240.2	1.3
Mortar	Small	2104976.0	14139229.1	0.8
Mortar	Small	2104976.7	14139229.3	0.9
Mortar	Small	2104013.6	14138535.2	1.1
Mortar	Small	2104017.5	14138533.9	0.9
NonOrd	Large	2105014.7	14139251.1	0.9
NonOrd	Large	2105139.8	14139037.1	1.1
[a] All dimensions are in feet.				

Demonstrator Declarations Not Within 2 m of a Baseline Item	
7[a]	No known baseline item within 2 m of this location.
13	No known baseline item within 2 m of this location.
176	No known baseline item within 2 m of this location.
[a] Target numbers are from the demonstrator numbering scheme.	

III. SUMMARY

The purpose of the 1994 Jefferson Proving Ground demonstration was to evaluate the state of the art in UXO detection and remediation technology. For all detection demonstrators, the maximum detection capabilities as measured by the target matching algorithm have been lower than $P_{match} = 0.65$ and there have been multiple false declarations per ordnance item detected. Further, statistical analysis of the JPG data indicates that demon-

strators with similar detection capabilities cannot be ordinally ranked with confidence. Therefore, demonstrators are binned on the basis of P_{match}, false alarm rate, and a separation criteria d, which combines measures of P_{match} and P_{fa} into a single variable and allows for comparison of demonstrators with widely different detection capabilities and false alarm rates. In this assessment, Geometrics performed in the vicinity of the average JPG demonstrator in terms of P_{match}, in the top bin of four bins in false alarms, and in the second of four bins in the separation measure.

This page is intentionally left blank

EVALUATION OF GEORADAR, INC.
40-ACRE SITE

I. DEMONSTRATOR RESULTS

	Entire Site	Area Searched[1]
Detection Capability		
overall detection ratio, R_{all}	0.004	0.07
ordnance detection ratio, R_{ord}	0.006	0.08
nonordnance detection ratio, R_{nonord}	0.0	0.0
mistyped ordnance ratio, *MR*	0.0	0.0
False Positive Ratio, *FPR*	*	*
False Negative Ratio, *FNR*	0.96	0.95
False Alarm Rate, *FAR*	0.13	2.6
Radial Accuracy, ΔR_{xy} *(feet)*	6.2	6.2
Depth Accuracy, ΔR_z *(feet)*	0.9	0.9
Target Classification Capabilities		
single targets	1	1
multiple targets	*	*
bombs	*	*
projectiles	0	0
mortars	*	*
mines	*	*
clusters	*	*

[1] The portion of the site searched by this demonstrator may not be representative of the overall difficulty of the entire 40-acre site.

* Ratio is undefined (i.e., zero denominator). A classification ratio is undefined if no targets of that class are located by the demonstrator.

Survey Rate

The demonstrator searched approximately 2 acres of the 40-acre site in the allotted 1 week. The proposal gave survey rates as 10 acres per day for hand-held magnetometer and 20–40 acres per day for the surface towed magnetometer. The demonstrator report

does not comment on the reason that survey rate fell short of the demonstrator's initial expectations.

II. DISCUSSION OF DEMONSTRATOR PERFORMANCE

GeoRadar, Inc. used a preproduction model of the GeoRadar 1000A man-portable ground penetrating radar sensor with a range of 5 to 10 feet for detection. Navigation was accomplished using markers placed at 10-foot intervals. The demonstrator proposal notes that ground penetrating radars have difficulty in wet, clay soils.

This demonstrator did not attempt to determine whether items were ordnance or nonordnance, typing all detections as ordnance items. Because no attempt was made to distinguish ordnance from nonordnance, several of the ratios used in this evaluation are not useful for characterizing this demonstrator. For example, $MR = 0$ not because the demonstrator made no errors in separating ordnance from nonordnance; rather, because no items were typed as nonordnance, no ordnance items could be misidentified as nonordnance. Of particular note for this demonstrator is the false negative ratio of 0.96. As a practical matter, if all the ordnance declarations of this demonstrator were investigated, some of the holes dug would contain emplaced nonordnance items and 96% of the remaining holes dug for remediation would not contain any of the ordnance items known by the government to be present. Further, less than 1% of the ordnance items would be remediated.

Corrections for random hits by this demonstrator reduced the uncorrected $P_{match} = 0.08$ to a corrected $P_{match} = 0.05$, where P_{match} refers to ordnance detection on only the area searched.

Detection capabilities were calculated for the various ordnance types emplaced. Only the targets in the area searched were considered, and credit was given for a detection regardless of classification ability, when calculating the detection capabilities reported in this paragraph. This demonstrator's best performance was in the detection of projectiles, with $P_{match} = 0.33$. This demonstrator had zero detections for all other target types. The performance of this demonstrator was best at intermediate depths. When the targets are grouped as those buried less than 3 feet, between 3 and 6 feet, and more than 6 feet deep, P_{match} was 0.0, 0.02, and 0.0, respectively. The demonstrator was best able to detect medium targets, with $P_{match} = 0.33$, and detected small and large targets both at $P_{match} = 0.0$.

The detection capabilities calculated above are all based on P_{match}, the probability of a one-to-one match between a demonstrator declaration and an emplaced target. This

method of determining detection probability assumes a high degree of sensor resolution, which may not be justified if a sensor cannot resolve signals resulting from two (or more) closely spaced objects. Two alternative measures of detection capability are also considered. For demonstrator declarations with multiple baseline items within R_{crit}, P_{group} credits the demonstrator with detecting a single group of targets, and does not penalize for any missed detections. P_{near} measures the probability that an item will be found if holes of radius R_{crit} are dug on all demonstrator declarations, allowing multiple baseline targets to be matched to a single demonstrator declaration. The following table compares detection capability calculated using these three measures, where P_{group} and P_{match} represent upper and lower bounds on the probability of detection, P_d, achieved on this demonstration. The false alarm rate, *FAR*, will be the same for all three measures; the false negative ratio will be different. Also presented in the table are P_{match} and P_{near} calculated with the plastic mines removed from the baseline. All detection measures are for ordnance items only.

Calculation method	Value
P_{match}[1]	0.08
P_{match}*(no mines)*	0.08
P_{group}	0.08
P_{near}	0.08
P_{near} *(no mines)*	0.08

[1] For this demonstration, the correction to detection capabilities for random matches was small (< 0.03).

The following tables provide information about specific demonstrator declarations and baseline items which may be of use to the demonstrator in assessing target identification capabilities for future operations, or for making improvements in the sensor or the data processing algorithms. The correct locations and types are provided for 17 baseline items that were dug up during or shortly after the remediation demonstrations. Five of these baseline items are isolated, individual objects. The remaining 12 are six closely spaced pairs of items. These items have either been removed from the site or were emplaced again at new locations. Three demonstrator declarations that were not within 2 meters of a baseline item are also provided.

Locations and Types of Selected Baseline Targets				
Type	Size	X[a]	Y	Depth
Bomb	Large	2104200.6	14139531.2	8.0
Bomb	Large	2105272.6	14139550.2	5.6
Bomb	Large	2105272.5	14139555.1	2.9
Projectile	Medium	2105087.4	14139006.9	3.4
Projectile	Medium	2105105.8	14139336.8	2.7
Projectile	Small	2105134.5	14139561.0	6.1
Projectile	Medium	2105326.1	14139555.1	2.0
Projectile	Medium	2105014.3	14139251.5	3.3
Mortar	Small	2105139.5	14139036.9	3.1
Mortar	Small	2105327.9	14139554.6	1.2
Mortar	Small	2104978.4	14139240.2	1.3
Mortar	Small	2104976.0	14139229.1	0.8
Mortar	Small	2104976.7	14139229.3	0.9
Mortar	Small	2104013.6	14138535.2	1.1
Mortar	Small	2104017.5	14138533.9	0.9
NonOrd	Large	2105014.7	14139251.1	0.9
NonOrd	Large	2105139.8	14139037.1	1.1
[a] All dimensions are in feet.				

Demonstrator Declarations Not Within 2 m of a Baseline Item	
2[a]	No known baseline item within 2 m of this location.
15	No known baseline item within 2 m of this location.
20	No known baseline item within 2 m of this location.
[a] Target numbers are from the demonstrator numbering scheme.	

III. SUMMARY

The purpose of the 1994 Jefferson Proving Ground demonstration was to evaluate the state of the art in UXO detection and remediation technology. For all detection demonstrators, the maximum detection capabilities as measured by the target matching algorithm have been lower than $P_{match} = 0.65$ and there have been multiple false declarations per ordnance item detected. Further, statistical analysis of the JPG data indicates that

demonstrators with similar detection capabilities cannot be ordinally ranked with confidence. Therefore, demonstrators are binned on the basis of P_{match}, false alarm rate, and a separation criteria d, which combines measures of P_{match} and P_{fa} into a single variable and allows for comparison of demonstrators with widely different detection capabilities and false alarm rates. In this assessment, GeoRadar performed significantly worse than the average JPG demonstrator in terms of P_{match}, in the third of four bins in false alarms, and in the last of four bins in the separation measure.

This page is intentionally left blank

EVALUATION OF JAYCOR
40-ACRE SITE

I. DEMONSTRATOR RESULTS

	Entire Site	Area Searched[1]
Detection Capability		
overall detection ratio, R_{all}	0.004	0.0‡
ordnance detection ratio, R_{ord}	0.006	0.0‡
nonordnance detection ratio, R_{nonord}	0.0	0.0
mistyped ordnance ratio, *MR*	0.0	*
False Positive Ratio, *FPR*	*	*
False Negative Ratio, *FNR*	0.99	1.0
False Alarm Rate, *FAR*	0.46	0.81
Radial Accuracy, ΔR_{xy} *(feet)*	4.3	*
Depth Accuracy, ΔR_z *(feet)*	**	**
Target Classification Capabilities		
single targets	1.0	*
multiple targets	*	*
bombs	0	*
projectiles	*	*
mortars	*	*
mines	*	*
clusters	*	*

[1] The portion of the site searched by this demonstrator may not be representative of the overall difficulty of the entire 40-acre site.

‡ The detection capability is lower on only the area searched because some of the declarations were outside the portion of the site that the demonstrator reported visiting.

* Ratio is undefined (i.e., zero denominator). A classification ratio is undefined if no targets of that class are located by the demonstrator.

** No depth information provided by demonstrator.

Survey Rate

The demonstrator searched approximately 20 acres of the 40-acre site within the allotted 1 week. The proposal states that four complete days are required to survey a 10-acre quadrant of the site. The demonstrator stopped surveying at noon on Friday due to an incoming storm.

II. DISCUSSION OF DEMONSTRATOR PERFORMANCE

Jaycor used two ground penetrating radar sensors mounted on a golf cart for detection. Navigation was accomplished by surveying the existing markers. Laboratory tests of the GPR system have demonstrated a 90% probability of detection for surface and buried metallic mines.

This demonstrator did not attempt to determine whether items were ordnance or nonordnance, typing all detections as ordnance items. Because no attempt was made to distinguish ordnance from nonordnance, several of the ratios used in this evaluation are not useful for characterizing this demonstrator. For example, $MR = 0$ not because the demonstrator made no errors in separating ordnance from nonordnance; rather, because no items were typed as nonordnance, no ordnance items could be misidentified as nonordnance. Of particular note for this demonstrator is the false negative ratio of 0.99. As a practical matter, if all the ordnance declarations of this demonstrator were investigated, some of the holes dug would contain emplaced nonordnance items and 99% of the remaining holes dug for remediation not contain any of the ordnance items known by the government to be present.

The demonstrator surveyed approximately 20 acres of the 40-acre site. When the demonstrator's detection capability and false alarm rates are recalculated to include only the portion of the site searched, the results are as follows:

$$P_{match} = 0.0$$

$$P_{match}\,(\text{ord}) = 0.0$$

$$FAR = 0.81$$

The P_{match} decreases compared to that calculated for the entire site, because the single target located by the demonstrator was in a grid cell that was not among those the demonstrator reported visiting.

Corrections for random hits by this demonstrator reduced the uncorrected P_{match} = 0.0 to a corrected $P_{match} = -0.011$, where P_{match} refers to the ordnance detection capability on only the area searched.

This demonstrator found no ordnance items, making the classification ratios and measures of detection capability in terms of target class, size, and depth meaningless.

The detection capabilities calculated above are all based on P_{match}, the probability of a one-to-one match between a demonstrator declaration and an emplaced target. This method of determining detection probability assumes a high degree of sensor resolution, which may not be justified if a sensor cannot resolve signals resulting from two (or more) closely spaced objects. Two alternative measures of detection capability are also considered. For demonstrator declarations with multiple baseline items within R_{crit}, P_{group} credits the demonstrator with detecting a single group of targets, and does not penalize for any missed detections. P_{near} measures the probability that an item will be found if holes of radius R_{crit} are dug on all demonstrator declarations, allowing multiple baseline targets to be matched to a single demonstrator declaration. The following table compares detection capability calculated using these three measures, where P_{group} and P_{match} represent upper and lower bounds on the probability of detection, P_d, achieved on this demonstration. The false alarm rate, *FAR*, will be the same for all three measures; the false negative ratio will be different. Also presented in the table are P_{match} and P_{near} calculated with the plastic mines removed from the baseline. All detection measures are for ordnance items only.

Calculation method	Value
P_{match}[1]	0.0
P_{match}(no mines)	0.0
P_{group}	0.0
P_{near}	0.0
P_{near} (no mines)	0.0

[1] For this demonstration, the correction to detection capabilities for random matches was small (< 0.03).

The following tables provide information about specific demonstrator declarations and baseline items which may be of use to the demonstrator in assessing target identification capabilities for future operations, or for making improvements in the sensor or the data processing algorithms. The correct locations and types are provided for 17 baseline items that were dug up during or shortly after the remediation demonstrations. Five of these baseline items are isolated, individual objects. The remaining 12 are six closely

spaced pairs of items. These items have either been removed from the site or were emplaced again at new locations. Three demonstrator declarations that were not within 2 meters of a baseline item are also provided.

Locations and Types of Selected Baseline Targets				
Type	Size	X[a]	Y	Depth
Bomb	Large	2104200.6	14139531.2	8.0
Bomb	Large	2105272.6	14139550.2	5.6
Bomb	Large	2105272.5	14139555.1	2.9
Projectile	Medium	2105087.4	14139006.9	3.4
Projectile	Medium	2105105.8	14139336.8	2.7
Projectile	Small	2105134.5	14139561.0	6.1
Projectile	Medium	2105326.1	14139555.1	2.0
Projectile	Medium	2105014.3	14139251.5	3.3
Mortar	Small	2105139.5	14139036.9	3.1
Mortar	Small	2105327.9	14139554.6	1.2
Mortar	Small	2104978.4	14139240.2	1.3
Mortar	Small	2104976.0	14139229.1	0.8
Mortar	Small	2104976.7	14139229.3	0.9
Mortar	Small	2104013.6	14138535.2	1.1
Mortar	Small	2104017.5	14138533.9	0.9
NonOrd	Large	2105014.7	14139251.1	0.9
NonOrd	Large	2105139.8	14139037.1	1.1

[a] All dimensions are in feet.

Demonstrator Declarations Not Within 2 m of a Baseline Item	
1[a]	No known baseline item within 2 m of this location.
10	No known baseline item within 2 m of this location.
75	No known baseline item within 2 m of this location.

[a] Target numbers are from the demonstrator numbering scheme.

III. SUMMARY

The purpose of the 1994 Jefferson Proving Ground demonstration was to evaluate the state of the art in UXO detection and remediation technology. For all detection

demonstrators, the maximum detection capabilities as measured by the target matching algorithm have been lower than $P_{match} = 0.65$ and there have been multiple false declarations per ordnance item detected. Further, statistical analysis of the JPG data indicates that demonstrators with similar detection capabilities cannot be ordinally ranked with confidence. Therefore, demonstrators are binned on the basis of P_{match}, false alarm rate, and a separation criteria d, which combines measures of P_{match} and P_{fa} into a single variable and allows for comparison of demonstrators with widely different detection capabilities and false alarm rates. In this assessment, Jaycor performed significantly worse than the average JPG demonstrator in terms of P_{match}, in the second of four bins in false alarms, and in the last of four bins in the separation measure.

This page is intentionally left blank

EVALUATION OF METRATEK, INC.
40-ACRE SITE

I. DEMONSTRATOR RESULTS

	Entire Site	Area Searched[1]
Detection Capability		
overall detection ratio, R_{all}	0.02	0.24
ordnance detection ratio, R_{ord}	0.03	0.31
nonordnance detection ratio, R_{nonord}	0.01	0.11
mistyped ordnance ratio, MR	0.0	0.0
False Positive Ratio, FPR	1.0	1.0
False Negative Ratio, FNR	0.90	0.89
False Alarm Rate, FAR	0.25	1.95
Radial Accuracy, ΔR_{xy} *(feet)*	2.9	2.9
Depth Accuracy, ΔR_z *(feet)*	**	**
Target Classification Capabilities		
single targets	0.8	0.8
multiple targets	*	*
bombs	0	0
projectiles	0	0
mortars	0	0
mines	*	*
clusters	*	*

[1] The portion of the site searched by this demonstrator may not be representative of the overall difficulty of the entire 40-acre site.

* Ratio is undefined (i.e., zero denominator). A classification ratio is undefined if no targets of that class are located by the demonstrator.

** No depth information provided.

Survey Rate

The demonstrator searched approximately 5 acres of the 40-acre site within the allotted 1 week. Because of equipment failures, the actual operating time of the GPR system was less than 2 full days. The induction coil did not experience any failures.

II. DISCUSSION OF DEMONSTRATOR PERFORMANCE

Metratek, Inc. used a prototype Model 200 stepped-frequency ground penetrating radar mounted on a sled pulled by a four-wheel-drive vehicle, and a man-portable Geonics EM61 induction coil metal detector. The prototype system used for this demonstration had a 4-foot swath and a frequency band of 0.2-0.65 GHz. The completed system is anticipated to have a 12-foot swath and a 0.2–2.0 GHz frequency band. Differential GPS was used for navigation. The demonstrator report noted that the ground conditions were muddy in low lying areas on Monday, and dried somewhat through Wednesday. It further noted that the conductivity of the soil was quite high, 30–70 millimhos/m at 40 MHz, which translated to losses on the order of 15–30 dB per foot in the low frequency ranges used and 40–80 dB per foot for the higher ranges, severely limiting detection of deep targets. The declarations of this demonstrator contained target identification numbers that were used more than once, making the input data incompatible with the target matching algorithm. Therefore, the target identification numbers of this demonstrator were changed to eliminate duplications prior to analysis.

The demonstrator surveyed approximately 5 acres of the 40-acre site. When the demonstrator's detection capability and false alarm rates are recalculated to include only the portion of the site searched, the results are as follows:

$$P_{match} = 0.24$$

$$P_{match}(\text{ord}) = 0.31$$

$$FAR = 1.95$$

Corrections for random hits by this demonstrator reduced the uncorrected P_{match} = 0.31 to a corrected P_{match} = 0.29, where P_{match} refers to the ordnance detection capability on only the area searched.

This demonstrator separated the targets detected using the GPR sensor from those detected using the electromagnetic sensor. The results for the two sensors, which follow, indicate that the majority of the target detections are attributable to the induction coil (IC) rather than the GPR.

	GPR	IC
P_{match}	0.04	0.20
$P_{match}(ord)$	0.06	0.25
FAR	0.59	1.4

Detection capabilities were calculated for the various ordnance types emplaced. Only the targets in the area searched were considered and credit was given for a detection regardless of classification ability, when calculating the detection capabilities reported in this paragraph. This demonstrator's best performance was in detection of projectiles, with $P_{match} = 0.43$. This demonstrator's worst performance was in detecting mortars, with $P_{match} = 0.20$. The portion of the site searched by this demonstrator did not contain any mines or cluster targets, which have had the lowest value of P_{match} among other demonstrators. The performance of this demonstrator was best at shallow and intermediate depths. When the targets are grouped as those buried less than 3 feet, between 3 and 6 feet, and more than 6 feet deep, P_{match} was 0.26, 0.33, and 0.0, respectively. The demonstrator was best able to detect medium targets, with $P_{match} = 0.43$, and detected small and large targets at $P_{match} =0.11$ and $P_{match} = 0.29$, respectively.

The detection capabilities calculated above are all based on P_{match}, the probability of a one-to-one match between a demonstrator declaration and an emplaced target. This method of determining detection probability assumes a high degree of sensor resolution, which may not be justified if a sensor cannot resolve signals resulting from two (or more) closely spaced objects. Two alternative measures of detection capability are also considered. For demonstrator declarations with multiple baseline items within R_{crit}, P_{group} credits the demonstrator with detecting a single group of targets, and does not penalize for any missed detections. P_{near} measures the probability that an item will be found if holes of radius R_{crit} are dug on all demonstrator declarations, allowing multiple baseline targets to be matched to a single demonstrator declaration. The following table compares detection capability calculated using these three measures, where P_{group} and P_{match} represent upper and lower bounds on the probability of detection, P_d, achieved on this demonstration. The false alarm rate, FAR, will be the same for all three measures; the false negative ratio will be different. Also presented in the table are P_{match} and P_{near} calculated with the plastic mines removed from the baseline. All detection measures are for ordnance items only.

Calculation method	Value
P_{match}[1]	0.31
$P_{match}(no\ mines)$	0.32
P_{group}	0.36
P_{near}	0.44
$P_{near}\ (no\ mines)$	0.44

1 For this demonstration, the correction to detection capabilities for random matches was small (< 0.03).

The following tables provide information about specific demonstrator declarations and baseline items which may be of use to the demonstrator in assessing target identification capabilities for future operations, or for making improvements in the sensor or the data processing algorithms. The correct locations and types are provided for 17 baseline items that were dug up during or shortly after the remediation demonstrations. Five of these baseline items are isolated, individual objects. The remaining 12 are six closely spaced pairs of items. These items have either been removed from the site or were emplaced again at new locations. Three demonstrator declarations that were not within 2 meters of a baseline item are also provided.

Locations and Types of Selected Baseline Targets				
Type	Size	X[a]	Y	Depth
Bomb	Large	2104200.6	14139531.2	8.0
Bomb	Large	2105272.6	14139550.2	5.6
Bomb	Large	2105272.5	14139555.1	2.9
Projectile	Medium	2105087.4	14139006.9	3.4
Projectile	Medium	2105105.8	14139336.8	2.7
Projectile	Small	2105134.5	14139561.0	6.1
Projectile	Medium	2105326.1	14139555.1	2.0
Projectile	Medium	2105014.3	14139251.5	3.3
Mortar	Small	2105139.5	14139036.9	3.1
Mortar	Small	2105327.9	14139554.6	1.2
Mortar	Small	2104978.4	14139240.2	1.3
Mortar	Small	2104976.0	14139229.1	0.8
Mortar	Small	2104976.7	14139229.3	0.9
Mortar	Small	2104013.6	14138535.2	1.1
Mortar	Small	2104017.5	14138533.9	0.9
NonOrd	Large	2105014.7	14139251.1	0.9
NonOrd	Large	2105139.8	14139037.1	1.1
[a] All dimensions are in feet.				

Demonstrator Declarations Not Within 2 m of a Baseline Item	
9[a]	No known baseline item within 2 m of this location.
41	No known baseline item within 2 m of this location.
47	No known baseline item within 2 m of this location.
[a] Target numbers are from the demonstrator numbering scheme.	

III. SUMMARY

The purpose of the 1994 Jefferson Proving Ground demonstration was to evaluate the state of the art in UXO detection and remediation technology. For all detection demonstrators, the maximum detection capabilities as measured by the target matching algorithm have been lower than $P_{match} = 0.65$ and there have been multiple false declarations per ordnance item detected. Further, statistical analysis of the JPG data indicates that

demonstrators with similar detection capabilities cannot be ordinally ranked with confidence. Therefore, demonstrators are binned on the basis of P_{match}, false alarm rate, and a separation criteria d, which combines measures of P_{match} and P_{fa} into a single variable and allows for comparison of demonstrators with widely different detection capabilities and false alarm rates. In this assessment, Metratek performed significantly better than the average JPG demonstrator in terms of P_{match}, in the second of four bins in false alarms, and in the third of four bins in the separation measure.

EVALUATION OF SECURITY SEARCH PRODUCTS/ VALLON GmbH 40-ACRE SITE

I. DEMONSTRATOR RESULTS

	Entire Site	Area Searched[1]
Detection Capability		
overall detection ratio, R_{all}	0.14	0.65
ordnance detection ratio, R_{ord}	0.11	0.62
nonordnance detection ratio, R_{nonord}	0.22	0.68
mistyped ordnance ratio, MR	0.00‡	0.00‡
False Positive Ratio, FPR	1.00‡	1.00‡
False Negative Ratio, FNR	0.98	0.98
False Alarm Rate, FAR	4.13	13.85
Radial Accuracy, ΔR_{xy} *(feet)*	2.0	2.0
Depth Accuracy, ΔR_z *(feet)*	1.4	1.5
Target Classification Capabilities		
single targets	1.0	1.0
multiple targets	0.0	0.0
bombs	0.60	0.60
projectiles	1.0	1.0
mortars	0.0	0.0
mines	*	*
clusters	0.0	0.0

[1] The portion of the site searched by this demonstrator may not be representative of the overall difficulty of the entire 40-acre site.

* Ratio is undefined (i.e., zero denominator). A classification ratio is undefined if no targets of that class are located by the demonstrator.

‡ The values shown here are a consequence of the demonstrator declaring all items as ordnance.

Survey Rate

The demonstrator searched approximately 12 acres of the 40-acre site within the allotted 1 week. The proposal stated that the 40-acre site would be surveyed in 4 days. The demonstrator report does not indicate the reason the survey rate fell short of the demonstrator's initial expectations.

II. DISCUSSION OF DEMONSTRATOR PERFORMANCE

Vallon proposed to use an array of five magnetometers towed by an all terrain vehicle to survey the site, with the option of removing a magnetometer for hand-held operation in areas where towing was impractical. Both hand-held and towed magnetometer and gradiometer systems were used. A SEPOS rope/tape/odometer system was used for navigation. The proposal states that objects buried up to 7 meters deep can be detected and that the accuracy of detection is 5 cm.

This demonstrator did not attempt to determine whether items were ordnance or nonordnance, typing all detections as ordnance items. Because no attempt was made to distinguish ordnance from nonordnance, several of the ratios used in this evaluation are not useful for characterizing this demonstrator. For example, $MR = 0$ not because the demonstrator made no errors in separating ordnance from nonordnance; rather, because no items were typed as nonordnance, no ordnance items could be misidentified as nonordnance. $FPR = 1.0$ since any nonordnance items found were typed as ordnance. Of particular note for this demonstrator is the false negative ratio of 0.98. As a practical matter, if all the ordnance declarations of this demonstrator were investigated, some of the holes dug would contain emplaced nonordnance items, and 98% of the remaining holes dug for remediation would not contain any of the targets known by the government to be present.

This demonstrator surveyed approximately 12 acres of the 40-acre site. When the demonstrator's detection capability and false alarm rate are recalculated to include only the portion of the site searched, the results are as follows:

$$P_{match} = 0.65$$

$$P_{match}(\text{ord}) = 0.62$$

$$FAR = 13.98$$

The FPR, FNR, and the classification ratios do not change. Although the demonstrator's P_{match} goes up substantially, the false alarm rating goes up significantly as well.

This demonstrator had the highest P_{match} on area searched, but searched only 12 acres of the 40-acre site, raising the issue of whether this portion of the site is representative, relatively easy, or relatively hard in comparison to the entire site. Therefore, we selected demonstrators whose overall P_{match} was comparable to Vallon's and compared their performance on the same 12-acre portion of the site that Vallon searched. The four demonstrators with the highest P_{match} on this portion of the site, including Vallon, scored 0.62, 0.59, 0.48, and 0.45. Of these demonstrators, Vallon scored the highest, but the differences in detection capability are much smaller when measured on only the portion of the site searched by Vallon, indicating that this area may be "easier" than the site as a whole or areas visited by other demonstrators. Finally, because Vallon searched only a small portion of the site, which contained only a small percentage of the emplaced objects, there is a large statistical uncertainty associated with the calculation of detection capability. There is no corresponding statistical uncertainty associated with this demonstrator's large false alarm rate.

Corrections for random hits by this demonstrator reduced the uncorrected P_{match} = 0.62 to a corrected P_{match} = 0.53, where P_{match} refers to the ordnance detection capability.

A number of the declarations by this demonstrator were indicated as pipe or fence in the comment field. When these declarations are removed, the results are as follows:

$$P_{match} = 0.63$$

$$P_{match}(\text{ord}) = 0.62$$

$$P_{match}(\text{corr}) = 0.55$$

$$FAR = 11.7$$

This correction does not substantially change P_{match}. The false alarm rate drops from 13.98 to 11.7, but this number is still much greater than the average false alarm rate of demonstrators on the 40-acre site and is exceeded by only one other demonstrator. The demonstrator with the next highest false alarm rate reported approximately a factor of two fewer false alarms per area visited.

Detection capabilities were calculated for the various ordnance types emplaced. Only the targets in the area searched were considered and credit was given for a detection regardless of classification ability, when calculating the detection capabilities reported in this paragraph. This demonstrator's best performance was in the detection of bombs and projectiles, with P_{match} = 0.83 and 0.78. Detection ability for clusters and mortars were

P_{match} = 0.50 and 0.42, respectively. The area visited by this demonstrator did not contain any mines, which have presented the greatest detection challenge to other demonstrators. The performance of this demonstrator was best at the lowest depths. When the targets are grouped as those buried less than 3 feet, between 3 and 6 feet, and more than 6 feet deep, P_{match} was 0.61, 0.59, and 1.0, respectively. The demonstrator was best able to detect medium targets, with P_{match} = 0.90, and detected small and large targets at P_{match} = 0.60 and P_{match} = 0.53, respectively.

The detection capabilities calculated above are all based on P_{match}, the probability of a one-to-one match between a demonstrator declaration and an emplaced target. This method of determining detection probability assumes a high degree of sensor resolution, which may not be justified if a sensor cannot resolve signals resulting from two (or more) closely spaced objects. Two alternative measures of detection capability are also considered. For demonstrator declarations with multiple baseline items within R_{crit}, P_{group} credits the demonstrator with detecting a single group of targets, and does not penalize for any missed detections. P_{near} measures the probability that an item will be found if holes of radius R_{crit} are dug on all demonstrator declarations, allowing multiple baseline targets to be matched to a single demonstrator declaration. The following table compares detection capability calculated using these three measures, where P_{group} and P_{match} represent upper and lower bounds on the probability of detection, P_d, achieved on this demonstration. The false alarm rate, *FAR*, will be the same for all three measures; the false negative ratio will be different. Also presented in the table are P_{match} and P_{near} calculated with the plastic mines removed from the baseline. All detection measures are for ordnance items only.

Calculation method	Value
P_{match}[1]	0.62
P_{match}(no mines)	0.62
P_{group}	0.72
P_{near}	0.76
P_{near} (no mines)	0.76

1 For this demonstrator, corrections for random matches reduced the P_{match} reported here to 0.55.

The following tables provide information about specific demonstrator declarations and baseline items which may be of use to the demonstrator in assessing target identification capabilities for future operations, or for making improvements in the sensor or the data processing algorithms. The correct locations and types are provided for 17

baseline items that were dug up during or shortly after the remediation demonstrations. Five of these baseline items are isolated, individual objects. The remaining 12 are six closely spaced pairs of items. These items have either been removed from the site or were emplaced again at new locations. Three demonstrator declarations that were not within 2 meters of a baseline item are also provided.

Locations and Types of Selected Baseline Targets				
Type	Size	X[a]	Y	Depth
Bomb	Large	2104200.6	14139531.2	8.0
Bomb	Large	2105272.6	14139550.2	5.6
Bomb	Large	2105272.5	14139555.1	2.9
Projectile	Medium	2105087.4	14139006.9	3.4
Projectile	Medium	2105105.8	14139336.8	2.7
Projectile	Small	2105134.5	14139561.0	6.1
Projectile	Medium	2105326.1	14139555.1	2.0
Projectile	Medium	2105014.3	14139251.5	3.3
Mortar	Small	2105139.5	14139036.9	3.1
Mortar	Small	2105327.9	14139554.6	1.2
Mortar	Small	2104978.4	14139240.2	1.3
Mortar	Small	2104976.0	14139229.1	0.8
Mortar	Small	2104976.7	14139229.3	0.9
Mortar	Small	2104013.6	14138535.2	1.1
Mortar	Small	2104017.5	14138533.9	0.9
NonOrd	Large	2105014.7	14139251.1	0.9
NonOrd	Large	2105139.8	14139037.1	1.1
[a] All dimensions are in feet.				

Demonstrator Declarations Not Within 2 m of a Baseline Item	
25[a]	No known baseline item within 2 m of this location.
11	No known baseline item within 2 m of this location.
752	No known baseline item within 2 m of this location.
[a] Target numbers are from the demonstrator numbering scheme.	

III. SUMMARY

The purpose of the 1994 Jefferson Proving Ground demonstration was to evaluate the state of the art in UXO detection and remediation technology. For all detection demonstrators, the maximum detection capabilities as measured by the target matching algorithm have been lower than $P_{match} = 0.65$ and there have been multiple false declarations per ordnance item detected. Further, statistical analysis of the JPG data indicates that demonstrators with similar detection capabilities cannot be ordinally ranked with confidence. Therefore, demonstrators are binned on the basis of P_{match}, false alarm rate, and a separation criteria d, which combines measures of P_{match} and P_{fa} into a single variable and allows for comparison of demonstrators with widely different detection capabilities and false alarm rates. In this assessment, Vallon performed significantly better than the average JPG demonstrator in terms of P_{match}, in the last of four bins in false alarms, and in the third of four bins in the separation measure.

EVALUATION OF SRI INTERNATIONAL 40-ACRE SITE

I. DEMONSTRATOR RESULTS

	Entire Site	Area Searched[1]
Detection Capability		
overall detection ratio, R_{all}	0.004	0.01
ordnance detection ratio, R_{ord}	0.0	0.0
nonordnance detection ratio, R_{nonord}	0.01	0.04
mistyped ordnance ratio, *MR*	*	*
False Positive Ratio, *FPR*	0.0‡	0.0‡
False Negative Ratio, *FNR*	*‡	*‡
False Alarm Rate, *FAR*	0.7	1.95
Radial Accuracy, ΔR_{xy} *(feet)*	6.4	6.4
Depth Accuracy, ΔR_z *(feet)*	**	**
Target Classification Capabilities		
single targets	*	*
multiple targets	*	*
bombs	*	*
projectiles	*	*
mortars	*	*
mines	*	*
clusters	*	*

1 The portion of the site searched by this demonstrator may not be representative of the overall difficulty of the entire 40-acre site.

* Ratio is undefined (i.e., zero denominator). The mistyped ratio and false positive ratios are undefined because all targets are identified as nonordnance. A classification ratio is undefined if no targets of that class are located by the demonstrator.

‡ The values shown here are a consequence of the demonstrator declaring all items as nonordnance.

** No depth information was provided.

Survey Rate

The demonstrator searched approximately 13 acres of the 40-acre site within the allotted 1 week. The proposal estimates that the 40-acre site could be covered in 2 to 3 days. The demonstrator report does not indicate why the survey rate fell short of the demonstrator's initial expectations.

III. DISCUSSION OF DEMONSTRATOR RESULTS

SRI, International used a trailer-mounted ground penetrating radar detection system. Two horn antennas look downward about 30° below the horizon, covering approximately 100 feet in range and 60° in azimuth. Navigation was accomplished by placing stakes every 100 feet in both directions to use as guides for moving about the site. Precise positions were determined using differential GPS. The demonstrator report indicates that the resistivity, as measured by SRI when the team was on-site, would result in attenuation losses through the soil such that maximum penetration of the radar would be less than 2 m.

Because all detections were classified as nonordnance items, several of the ratios used in this evaluation are not useful for characterizing this demonstrator. Most seriously, the 0.0 value for the false positive ratio implies that this demonstrator has no such false declarations. However, this number indicates only that no items were declared ordnance. The false negative ratio is undefined because it requires that some items be declared as ordnance for a non-zero denominator. *MR* is undefined because this demonstrator did not detect any ordnance items.

SRI surveyed approximately one third of the site. When the demonstrator's overall detection capability and false alarm rate are recalculated using only the area searched, the results are as follows:

$$P_{match} = 0.011$$

$$P_{match}(\text{ord}) = 0.00$$

$$FAR = 1.95$$

Corrections for random hits by this demonstrator reduced the uncorrected P_{match} = 0.0 to a corrected $P_{match} = -0.03$, where P_{match} refers to the ordnance detection capability on only the area searched.

This demonstrator found no ordnance items, making the classification ratios and measures of detection capability in terms of target classification, size, and depth meaningless.

The detection capabilities calculated above are all based on P_{match}, the probability of a one-to-one match between a demonstrator declaration and an emplaced target. This method of determining detection probability assumes a high degree of sensor resolution, which may not be justified if a sensor cannot resolve signals resulting from two (or more) closely spaced objects. Two alternative measures of detection capability are also considered. For demonstrator declarations with multiple baseline items within R_{crit}, P_{group} credits the demonstrator with detecting a single group of targets, and does not penalize for any missed detections. P_{near} measures the probability that an item will be found if holes of radius R_{crit} are dug on all demonstrator declarations, allowing multiple baseline targets to be matched to a single demonstrator declaration. The following table compares detection capability calculated using these three measures, where P_{group} and P_{match} represent upper and lower bounds on the probability of detection, P_d, achieved on this demonstration. The false alarm rate, *FAR*, will be the same for all three measures; the false negative ratio will be different. Also presented in the table are P_{match} and P_{near} calculated with the plastic mines removed from the baseline. All detection measures are for ordnance items only.

Calculation method	Value
P_{match}[1]	0.0
P_{match}(no mines)	0.0
P_{group}	0.0
P_{near}	0.0
P_{near} (no mines)	0.0

[1] For this demonstration, the correction to detection capabilities for random matches was small (< 0.03).

The following tables provide information about specific demonstrator declarations and baseline items which may be of use to the demonstrator in assessing target identification capabilities for future operations, or for making improvements in the sensor or the data processing algorithms. The correct locations and types are provided for 17 baseline items that were dug up during or shortly after the remediation demonstrations. Five of these baseline items are isolated, individual objects. The remaining 12 are six closely spaced pairs of items. These items have either been removed from the site or were

emplaced again at new locations. Three demonstrator declarations that were not within 2 meters of a baseline item are also provided.

Locations and Types of Selected Baseline Targets				
Type	Size	X[a]	Y	Depth
Bomb	Large	2104200.6	14139531.2	8.0
Bomb	Large	2105272.6	14139550.2	5.6
Bomb	Large	2105272.5	14139555.1	2.9
Projectile	Medium	2105087.4	14139006.9	3.4
Projectile	Medium	2105105.8	14139336.8	2.7
Projectile	Small	2105134.5	14139561.0	6.1
Projectile	Medium	2105326.1	14139555.1	2.0
Projectile	Medium	2105014.3	14139251.5	3.3
Mortar	Small	2105139.5	14139036.9	3.1
Mortar	Small	2105327.9	14139554.6	1.2
Mortar	Small	2104978.4	14139240.2	1.3
Mortar	Small	2104976.0	14139229.1	0.8
Mortar	Small	2104976.7	14139229.3	0.9
Mortar	Small	2104013.6	14138535.2	1.1
Mortar	Small	2104017.5	14138533.9	0.9
NonOrd	Large	2105014.7	14139251.1	0.9
NonOrd	Large	2105139.8	14139037.1	1.1

[a] All dimensions are in feet.

Demonstrator Declarations Not Within 2 m of a Baseline Item	
2[a]	No known baseline item within 2 m of this location.
10	No known baseline item within 2 m of this location.
120	No known baseline item within 2 m of this location.

[a] Target numbers are from the demonstrator numbering scheme.

III. SUMMARY

The purpose of the 1994 Jefferson Proving Ground demonstration was to evaluate the state of the art in UXO detection and remediation technology. For all detection demonstrators, the maximum detection capabilities as measured by the target matching algorithm

have been lower than $P_{match} = 0.65$ and there have been multiple false declarations per ordnance item detected. Further, statistical analysis of the JPG data indicates that demonstrators with similar detection capabilities cannot be ordinally ranked with confidence. Therefore, demonstrators are binned on the basis of P_{match}, false alarm rate, and a separation criteria d, which combines measures of P_{match} and P_{fa} into a single variable and allows for comparison of demonstrators with widely different detection capabilities and false alarm rates. In this assessment, SRI performed significantly worse than the average JPG demonstrator in terms of P_{match}, in the second of four bins in false alarms, and in the last of four bins in the separation measure.

This page is intentionally left blank

EVALUATION OF UXB INTERNATIONAL, INC.
40-ACRE SITE

I. DEMONSTRATOR RESULTS

	Entire Site	Area Searched[1]
Detection Capability		
overall detection ratio, R_{all}	0.38	0.42
ordnance detection ratio, R_{ord}	0.33	0.36
nonordnance detection ratio, R_{nonord}	0.47	0.53
mistyped ordnance ratio, MR	1.0‡	1.0‡
False Positive Ratio, FPR	0.0‡	0.0‡
False Negative Ratio, FNR	*‡	*‡
False Alarm Rate, FAR	1.19	1.51‡
Radial Accuracy, ΔR_{xy} *(feet)*	1.91	1.87
Depth Accuracy, ΔR_z *(feet)*	1.74	1.73
Target Classification Capabilities		
single targets	0.0	0.0
multiple targets	*	*
bombs	0.0	0.0
projectiles	0.0	0.0
mortars	0.0	0.0
mines	*	*
clusters	*	*

[1] The portion of the site searched by this demonstrator may not be representative of the overall difficulty of the entire 40-acre site.

* Ratio is undefined (i.e., zero denominator). A classification ratio is undefined if no targets of that class are located by the demonstrator.

‡ The values shown here are a consequence of the demonstrator declaring all items as nonordnance.

Survey Rate

The demonstrator searched approximately 30 acres of the 40-acre site within the allotted 1 week. The proposal gave the survey rate as 3 to 4.5 acres per day (i.e., 15–22.5 acres per week).

II. DISCUSSION OF DEMONSTRATOR PERFORMANCE

UXB used hand-carried Schonstedt GA-52B and Foerster Ferex magnetometers to survey the site. The Global Position System was used for navigation. The UXB proposal indicates that the GA-52B sensors will work to a depth of 3 meters and the Ferex sensors to a depth of 19 ft (5.8 m). The proposal further states that the conditions at JPG "more than meet" the ideal conditions set forth by the company regarding terrain, vegetation, and ferrous debris.

Because all detections by this demonstrator were classified "other" and treated as nonordnance items, several of the ratios used in this evaluation are not useful for characterizing this demonstrator. Most seriously, the 0.0 value for the false positive ratio implies that this demonstrator has no such false declarations. However, these numbers indicate only that no items were declared as ordnance. *MR* has a value of 1.0, as all ordnance was mistyped (i.e., declared nonordnance). Similarly, the false negative ratio is undefined because it attempts to measure the fraction of ordnance declarations that will be false alarms, but the demonstrator had no ordnance declarations. All classification ratios are also zero for this reason, with the exception of the multiple target, mine, and cluster ratios, which are undefined because the demonstrator failed to locate any of these targets.

The demonstrator surveyed approximately 30 acres of the 40-acre site. When the demonstrator's detection capability and false alarm rates are recalculated to include only the portion of the site searched, the results are as follows:

$$P_{match} = 0.42$$

$$P_{match}(\text{ord}) = 0.36$$

$$FAR = 1.51$$

Corrections for random hits by this demonstrator reduced the uncorrected P_{match} = 0.36 to a corrected P_{match} = 0.35, where P_{match} refers to the ordnance detection capability on only the area searched.

The detection capabilities calculated above are all based on P_{match}, the probability of a one-to-one match between a demonstrator declaration and an emplaced target. This method of determining detection probability assumes a high degree of sensor resolution, which may not be justified if a sensor cannot resolve signals resulting from two (or more) closely spaced objects. Two alternative measures of detection capability are also considered. For demonstrator declarations with multiple baseline items within R_{crit}, P_{group} credits the demonstrator with detecting a single group of targets, and does not penalize for any missed detections. P_{near} measures the probability that an item will be found if holes of radius R_{crit} are dug on all demonstrator declarations, allowing multiple baseline targets to be matched to a single demonstrator declaration. The following table compares detection capability calculated using these three measures, where P_{group} and P_{match} represent upper and lower bounds on the probability of detection, P_d, achieved on this demonstration. The false alarm rate, *FAR*, will be the same for all three measures; the false negative ratio will be different. Also presented in the table are P_{match} and P_{near} calculated with the plastic mines removed from the baseline. All detection measures are for ordnance items only.

Calculation method	Value
P_{match}[1]	0.36
P_{match}(no mines)	0.43
P_{group}	0.44
P_{near}	0.55
P_{near} (no mines)	0.65

1 For this demonstration, the correction to detection capabilities for random matches was small (< 0.03).

Detection capabilities were calculated for the various ordnance types emplaced. Only the targets in the area searched were considered, and credit was given for a detection regardless of classification ability, when calculating the detection capabilities reported in this paragraph. This demonstrator's best performance was in the detection of bombs and projectiles, with $P_{match} = 0.58$ and $P_{match} = 0.53$, respectively. This demonstrator's worst performance was in detecting mines and cluster targets, with zero detections of either target type. Since the sensors used by this demonstrator can detect only metal objects, the inability to detect plastic mines is to be expected. The performance of this demonstrator was best at intermediate depths. When the targets are grouped as those buried less than 3 feet, between 3 and 6 feet, and more than 6 feet deep, P_{match} was 0.41, 0.47, and 0.35, respectively. The demonstrator was best able to detect medium size targets, with P_{match} = 0.59, and detected small and large targets at $P_{match} = 0.34$ and $P_{match} = 0.44$, respectively.

The following tables provide information about specific demonstrator declarations and baseline items which may be of use to the demonstrator in assessing target identification capabilities for future operations, or for making improvements in the sensor or the data processing algorithms. The correct locations and types are provided for 17 baseline items that were dug up during or shortly after the remediation demonstrations. Five of these baseline items are isolated, individual objects. The remaining 12 are six closely spaced pairs of items. These items have either been removed from the site or were emplaced again at new locations. Three demonstrator declarations that were not within 2 meters of a baseline item are also provided.

Locations and Types of Selected Baseline Targets				
Type	Size	X[a]	Y	Depth
Bomb	Large	2104200.6	14139531.2	8.0
Bomb	Large	2105272.6	14139550.2	5.6
Bomb	Large	2105272.5	14139555.1	2.9
Projectile	Medium	2105087.4	14139006.9	3.4
Projectile	Medium	2105105.8	14139336.8	2.7
Projectile	Small	2105134.5	14139561.0	6.1
Projectile	Medium	2105326.1	14139555.1	2.0
Projectile	Medium	2105014.3	14139251.5	3.3
Mortar	Small	2105139.5	14139036.9	3.1
Mortar	Small	2105327.9	14139554.6	1.2
Mortar	Small	2104978.4	14139240.2	1.3
Mortar	Small	2104976.0	14139229.1	0.8
Mortar	Small	2104976.7	14139229.3	0.9
Mortar	Small	2104013.6	14138535.2	1.1
Mortar	Small	2104017.5	14138533.9	0.9
NonOrd	Large	2105014.7	14139251.1	0.9
NonOrd	Large	2105139.8	14139037.1	1.1
[a] All dimensions are in feet.				

Demonstrator Declarations Not Within 2 m of a Baseline Item	
301[a]	No known baseline item within 2 m of this location.
305	No known baseline item within 2 m of this location.
310	No known baseline item within 2 m of this location.
[a] Target numbers are from the demonstrator numbering scheme.	

III. SUMMARY

The purpose of the 1994 Jefferson Proving Ground demonstration was to evaluate the state of the art in UXO detection and remediation technology. For all detection demonstrators, the maximum detection capabilities as measured by the target matching algorithm have been lower than $P_{match} = 0.65$ and there have been multiple false declarations per ordnance item detected. Further, statistical analysis of the JPG data indicates that demonstrators with similar detection capabilities cannot be ordinally ranked with confidence. Therefore, demonstrators are binned on the basis of P_{match}, false alarm rate, and a separation criteria d, which combines measures of P_{match} and P_{fa} into a single variable and allows for comparison of demonstrators with widely different detection capabilities and false alarm rates. In this assessment, UXB performed significantly better than the average JPG demonstrator in terms of P_{match}, in the second of four bins in false alarms, and in the second of four bins in the separation measure.

This page is intentionally left blank

6. 80-ACRE SITE EVALUATIONS

EVALUATION OF AIRBORNE ENVIRONMENTAL SURVEYS 80-ACRE SITE

I. DEMONSTRATOR RESULTS

	R_{crit} = 2 m	R_{crit} = 5 m	R_{crit} = 10 m
Detection Capability			
overall detection ratio	0.005	0.01	0.02
ordnance detection ratio	0.006	0.01	0.02
nonordnance detection ratio	0	0	0
mistyped ordnance ratio	0	0	0
False Positive Ratio, *FPR*	*	*	*
False Negative Ratio, *FNR*	0.97	0.95	0.89
False Alarm Rate, *FAR*	0.13	0.12	0.12
Radial Accuracy, ΔR_{xy}	6.0	7.7	18.4
Depth Accuracy, ΔR_z	0.94	0.91	3.96

* Ratio is undefined (i.e., zero denominator).

II. DISCUSSION OF DEMONSTRATOR PERFORMANCE

Airborne Environmental Surveys used ground penetrating radar and infrared imaging systems on a helicopter platform. Two wideband frequency-modulated radars were used; one was centered at 500 MHz and the other at 3 GHz. The infrared detector was a FLIR 2000F imager. Differential global positioning was used for navigation. A circular error of < 5 m was anticipated. The entire 80-acre site was visited during the allotted time.

The detection capabilities calculated above are based on the probability of a one-to-one match between a demonstrator declaration and an emplaced target. This method of determining detection probability assumes a high degree of sensor resolution, which may not be justified. If a sensor cannot resolve signals resulting from two (or more) closely spaced objects, this approach penalizes the demonstrator for the lack of resolution. A more generous measure of detection capability would be the probability that an item will be found if holes of radius R_{crit} are dug on all demonstrator declarations, allowing for the possibility

that multiple baseline targets may be found with a single demonstrator declaration. Since neither of these measures exactly corresponds to the probability of detection in the conventional sense, we refer to them as P_{match} and P_{near}.

In the absence of detailed sensor models for the demonstrator, we use R_{crit} as a surrogate for resolution and devise a third measure. This third measure does not count as missed targets any baseline items which are unmatched by the one-to-one algorithm but are within R_{crit} of a demonstrator's hit. In other words, multiple baseline items within R_{crit} of a demonstrator declaration are aggregated into a single baseline item. This measure, P_{group}, is by construction intermediate between the other two, and the one most likely to prove a reasonable surrogate for detection as it is usually meant. The following table compares P_d calculated using these three measures, where P_{group} and P_{match} represent upper and lower bounds on the probability of detection achieved on this demonstration. The false alarms will be the same for P_{match}, P_{near}, and P_{group}. Also presented in the table are P_{match} and P_{near} calculated with the plastic mines removed from the baseline. All P_ds are for ordnance items only.

The detection probabilities were generally so low for the 80-acre site that it is important to correct the P_{match} computed by the target matching algorithm for the number of matches that would have been made by the same number of demonstrator declarations placed randomly on the site. The corrected P_{match} values are essentially zero or negative. Coupled with the small absolute number of matches, this shows that there is no statistically significant evidence that matching of this demonstrator's reported declarations to emplaced ordnance items is other than a random occurrence. The table is included in part to show how small detection capability is by any measure. Even when more generous measures such as P_{group} or P_{near} are considered, or when a very large R_{crit} is used to allow for inaccurate location ability, detection capabilities are insignificant.

P_d calculation method	R_{crit} = 2 m	R_{crit} = 5 m	R_{crit} = 10 m
P_{match}*	0.01	0.01	0.02
P_{match} (no mines)	0.01	0.01	0.03
P_{group}	0.01	0.01	0.02
P_{near}	0.01	0.02	0.02
P_{near} (no mines)	0.01	0.02	0.03

* For this demonstrator, corrections for random matches reduced P_{match} to 0.0 for both R_{crit} = 2 m and 5 m. For R_{crit} = 10 m, the approximations made in the correction become invalid, making the correction term artifically large.

To further support this notion, if matches are generated by randomly placed hits, the expected location accuracy is given by $\Delta R_{xy} = 2/3\ R_{crit}$. These calculated values, shown in the table below, match well to the observed location accuracies of this demonstrator, further supporting the notion that no demonstrator declarations are traceable to an emplaced ordnance item.

	$R_{crit} = 2$ m	$R_{crit} = 5$ m	$R_{crit} = 10$ m
ΔR_{xy} *(ft)*	6.0	7.7	18.4
$2/3\ R_{crit}$ *(ft)*	4.4	10.9	21.9

The standard deviation also scales with R_{crit} and has approximately the expected value.

Given that there is no confidence in the matches resulting from the emplaced ordnance, the classification capabilities for the airborne demonstrators were not computed.

The following tables provide information about specific demonstrator declarations and baseline targets that may be of use to the demonstrator in assessing target identification capabilities for future operations, or for making improvements in the sensor or the data processing algorithms. The correct locations are provided for four mortars, three of which are on the surface. Three demonstrator declarations that were not within 5 m of a baseline item are also provided.

Locations and Types of Selected Baseline Targets				
Type	Size	X[a]	Y	Depth
Mortar	Small	2101548.4	14120689.8	2.06
Mortar	Small	2102121.0	14121733.6	0.0
Mortar	Small	2102061.1	14121777.9	0.0
Mortar	Small	2101976.8	14121862.1	0.0
[a] All dimensions are in feet. UTM Coordinates are U.S. Survey feet.				

Demonstrator Declarations Not Within 5 m of a Baseline Item	
7[a]	No known baseline item within 5 m of this location.
11	No known baseline item within 5 m of this location.
15	No known baseline item within 5 m of this location.
[a] Target numbers are from the demonstrator numbering scheme. Positions and types are correct for emplaced targets.	

III. SUMMARY

The purpose of the 1994 Jefferson Proving Ground demonstration was to evaluate the state of the art in UXO detection and remediation technology. For all airborne detection demonstrators, the detection capabilities have been statistically indistinguishable from zero.

This page is intentionally left blank

EVALUATION OF BATTELLE 80-ACRE SITE

I. DEMONSTRATOR RESULTS

	R_{crit} = 2 m	R_{crit} = 5 m	R_{crit} = 10 m
Detection Capability			
overall detection ratio	0	0	0.07
ordnance detection ratio	0	0	0.08
nonordnance detection ratio	0	0	0
mistyped ordnance ratio	*	*	0
False Positive Ratio, *FPR*	*	*	*
False Negative Ratio, *FNR*	1.0	1.0	0.92
False Alarm Rate, *FAR*	0.50	0.50	0.46
Radial Accuracy, ΔR_{xy}	*	*	26.6
Depth Accuracy, ΔR_z	*	*	2.3

* Ratio is undefined (i.e., zero denominator). For this demonstrator, these results are for the approximately 29 acres which contain all the demonstrator declarations. All other airborne demonstrators searched the entire site.

II. DISCUSSION OF DEMONSTRATOR PERFORMANCE

Battelle used a stepped-chirp radar system operating between 50 and 750 MHz, with the antennas looking 15° below the horizon. The system is intended for use on an airborne platform, but for the JPG demonstration it was mounted on the boom of a cherry-picker, which was driven along the roadway. Data were taken with the antennas at 40, 50, and 60 foot elevations. GPS was used for navigation. Battelle reported visiting almost the entire 80-acre site during the allotted time. However, all their hits are within 500 ft of the edge of the site. They were scored in detection and false alarm rate as if only the 29 acres of the region containing hits was effectively searched. We suspect that the radar when mounted on the cherry-picker is ineffective beyond a range of several hundred feet.

The detection capabilities calculated above are based on the probability of a one-to-one match between a demonstrator declaration and an emplaced target. This method of

determining detection probability assumes a high degree of sensor resolution, which may not be justified. If a sensor cannot resolve signals resulting from two (or more) closely spaced objects, this approach penalizes the demonstrator for the lack of resolution. A more generous measure of detection capability would be the probability that an item will be found if holes of radius R_{crit} are dug on all demonstrator declarations, allowing for the possibility that multiple baseline targets may be found with a single demonstrator declaration. Since neither of these measures exactly corresponds to the probability of detection in the conventional sense, we refer to them as P_{match} and P_{near}.

In the absence of detailed sensor models for the demonstrator, we use R_{crit} as a surrogate for resolution and devise a third measure. This third measure does not count as missed targets any baseline items which are unmatched by the one-to-one algorithm but are within R_{crit} of a demonstrator's hit. In other words, multiple baseline items within R_{crit} of a demonstrator declaration are aggregated into a single baseline item. This measure, P_{group}, is by construction intermediate between the other two, and the one most likely to prove a reasonable surrogate for detection as it is usually meant. The following table compares P_d calculated using these three measures, where P_{group} and P_{match} represent upper and lower bounds on the probability of detection achieved on this demonstration. The false alarms will be the same for P_{match}, P_{near}, and P_{group}. Also presented in the table are P_{match} and P_{near} calculated with the plastic mines removed from the baseline. All P_ds are for ordnance items only.

The detection probabilities were so low for the 80-acre site generally that it is important to correct the P_{match} computed by the target matching algorithm for the number of matches that would have been made by the same number of demonstrator declarations placed randomly on the site. The corrected P_{match} values are essentially zero or negative. Coupled with the small absolute number of matches, this shows that there is no statistically significant evidence that matching of this demonstrator's reported declarations to emplaced ordnance items is other than a random occurrence. The table is included in part to show how small detection capability is by any measure. Even when more generous measures such as P_{group} or P_{near} are considered, or when a very large R_{crit} is used to allow for inaccurate location ability, detection capabilities are insignificant.

P_d calculation method	R_{crit} = 2 m	R_{crit} = 5 m	R_{crit} = 10 m
P_{match}*	0.0	0.0	0.04
P_{match} (no mines)	0.0	0.0	0.03
P_{group}	0.0	0.0	0.04
P_{near}	0.0	0.0	0.07
P_{near} (no mines)	0.0	0.0	0.08

* For this demonstrator, corrections for random matches reduced P_{match} to 0.0 and – 0.03 for R_{crit} = 2 m and 5 m, respectively. For R_{crit} = 10 m, the approximations made in the correction become invalid, making the correction term artifically large.

To further support this notion, if matches are generated by randomly placed hits, the expected location accuracy is given by $\Delta R_{xy} = 2/3\ R_{crit}$. These calculated values, shown in the table below, match well to the observed location accuracies of this demonstrator, further supporting the notion that no demonstrator declarations are traceable to an emplaced ordnance item.

	R_{crit} = 2 m	R_{crit} = 5 m	R_{crit} = 10 m
ΔR_{xy} (ft)	-	-	26.6
2/3 R_{crit} (ft)	4.4	10.9	21.9

The standard deviation also scales with R_{crit} and has approximately the expected value.

Given that there is no confidence in the matches resulting from the emplaced ordnance, the classification capabilities for the airborne demonstrators were not computed.

The following tables provide information about specific demonstrator declarations and baseline targets that may be of use to the demonstrator in assessing target identification capabilities for future operations, or for making improvements in the sensor or the data processing algorithms. The correct locations are provided for four mortars, three of which are on the surface. Three demonstrator declarations that were not within 5 m of a baseline item are also provided.

Locations and Types of Selected Baseline Targets				
Type	Size	X[a]	Y	Depth
Mortar	Small	2101548.4	14120689.8	2.06
Mortar	Small	2102121.0	14121733.6	0.0
Mortar	Small	2102061.1	14121777.9	0.0
Mortar	Small	2101976.8	14121862.1	0.0
[a] All dimensions are in feet. UTM Coordinates are U.S. Survey feet.				

Demonstrator Declarations Not Within 5 m of a Baseline Item	
4[a]	No known baseline item within 5 m of this location.
11	No known baseline item within 5 m of this location.
29	No known baseline item within 5 m of this location.
[a] Target numbers are from the demonstrator numbering scheme. Positions and types are correct for emplaced targets.	

III. SUMMARY

The purpose of the 1994 Jefferson Proving Ground demonstration was to evaluate the state of the art in UXO detection and remediation technology. For all airborne detection demonstrators, the detection capabilities have been statistically indistinguishable from zero.

EVALUATION OF GEONEX AERODAT, INC.
80-ACRE SITE

I. DEMONSTRATOR RESULTS

	R_{crit} = 2 m	R_{crit} = 5 m	R_{crit} = 10 m
Detection Capability			
overall detection ratio	0	0.04	0.09
ordnance detection ratio	0	0.04	0.06
nonordnance detection ratio	0	0.08	0.38
mistyped ordnance ratio	*	0	0
False Positive Ratio, *FPR*	*	1.0	1.0
False Negative Ratio, *FNR*	1.0	0.95	0.93
False Alarm Rate, *FAR*	0.42	0.39	0.39
Radial Accuracy, ΔR_{xy}	*	10.3	17.4
Depth Accuracy, ΔR_z	**	**	**

* Ratio is undefined (i.e., zero denominator).

** No depth information provided.

II. DISCUSSION OF DEMONSTRATOR PERFORMANCE

Geonex/Aerodat used a magnetometer/gradiometer and electromagnetic induction sensor on a helicopter platform. Two cesium vapor magnetometers are mounted on opposite ends of a 6-m Kevlar tube towed below a helicopter. The electromagnetic transmitter and receiver are housed within the tube, along with position and attitude sensors. Laser, radar, and barometric altimeters measure the height of the helicopter, and differential GPS was to be used for navigation. The DGPS system failed, so navigation was limited to survey lanes marked by ground-based personnel. Positioning was further complicated by swaying and pendular effects experienced by the Kevlar bar during strong winds. The Geonex system experienced 31 hours of down-time due to weather and equipment failures. The entire 80-acre site was visited during the allotted time.

The detection capabilities calculated above are based on the probability of a one-to-one match between a demonstrator declaration and an emplaced target. This method of

determining detection probability assumes a high degree of sensor resolution, which may not be justified. If a sensor cannot resolve signals resulting from two (or more) closely spaced objects, this approach penalizes the demonstrator for the lack of resolution. A more generous measure of detection capability would be the probability that an item will be found if holes of radius R_{crit} are dug on all demonstrator declarations, allowing for the possibility that multiple baseline targets may be found with a single demonstrator declaration. Since neither of these measures exactly corresponds to the probability of detection in the conventional sense, we refer to them as P_{match} and P_{near}.

In the absence of detailed sensor models for the demonstrator, we use R_{crit} as a surrogate for resolution and devise a third measure. This third measure does not count as missed targets any baseline items which are unmatched by the one-to-one algorithm but are within R_{crit} of a demonstrator's hit. In other words, multiple baseline items within R_{crit} of a demonstrator declaration are aggregated into a single baseline item. This measure, P_{group}, is by construction intermediate between the other two, and the one most likely to prove a reasonable surrogate for detection as it is usually meant. The following table compares P_d calculated using these three measures, where P_{group} and P_{match} represent upper and lower bounds on the probability of detection achieved on this demonstration. The false alarms will be the same for P_{match}, P_{near}, and P_{group}. Also presented in the table are P_{match} and P_{near} calculated with the plastic mines removed from the baseline. All P_ds are for ordnance items only.

The detection probabilities were so low for the 80-acre site generally that it is important to correct the P_{match} computed by the target matching algorithm for the number of matches that would have been made by the same number of demonstrator declarations placed randomly on the site. The corrected P_{match} values are essentially zero or negative. Coupled with the small absolute number of matches, this shows that there is no statistically significant evidence that matching of this demonstrator's reported declarations to emplaced ordnance items is other than a random occurrence. The table is included in part to show how small detection capability is by any measure. Even when more generous measures such as P_{group} or P_{near} are considered, or when a very large R_{crit} is used to allow for inaccurate location ability, detection capabilities are insignificant.

P_d calculation method	R_{crit} = 2 m	R_{crit} = 5 m	R_{crit} = 10 m
P_{match}*	0.0	0.04	0.06
P_{match} (no mines)	0.0	0.05	0.06
P_{group}	0.0	0.04	0.06
P_{near}	0.0	0.06	0.17
P_{near} (no mines)	0.0	0.06	0.12

* For this demonstrator, corrections for random matches reduced P_{match} to – 0.01 and 0.01 for R_{crit} = 2 m and 5 m, respectively. For R_{crit} = 10 m, the approximations made in the correction become invalid, making the correction term artifically large.

To further support this notion, if matches are generated by randomly placed hits, the expected location accuracy is given by $\Delta R_{xy} = 2/3\ R_{crit}$. These calculated values, shown in the table below, match well to the observed location accuracies of this demonstrator, further supporting the notion that no demonstrator declarations are traceable to an emplaced ordnance item.

	R_{crit} = 2 m	R_{crit} = 5 m	R_{crit} = 10 m
ΔR_{xy} (ft)	-	10.3	17.4
2/3 R_{crit} (ft)	4.4	10.9	21.9

The standard deviation also scales with R_{crit} and has approximately the expected value.

Given that there is no confidence in the matches resulting from the emplaced ordnance, the classification capabilities for the airborne demonstrators were not computed.

The following table provides information about specific demonstrator declarations and baseline targets that may be of use to the demonstrator in assessing target identification capabilities for future operations, or for making improvements in the sensor or the data processing algorithms. The correct locations are provided for four mortars, three of which are on the surface. Three demonstrator declarations that were not within 5 m of a baseline item are also provided.

Location and Types of Selected Baseline Targets				
Type	Size	X[a]	Y	Depth
Mortar	Small	2101548.4	14120689.8	2.06
Mortar	Small	2102121.0	14121733.6	0.0
Mortar	Small	2102061.1	14121777.9	0.0
Mortar	Small	2101976.8	14121862.1	0.0

[a] All dimensions are in feet. UTM Coordinates are U.S. Survey feet.

Demonstrator Declarations Not Within 5 m of a Baseline Item	
1031[a]	No known baseline item within 5 m of this location.
1067	No known baseline item within 5 m of this location.
1133	No known baseline item within 5 m of this location.

[a] Target numbers are from the demonstrator numbering scheme. Positions and types are correct for emplaced targets.

III. SUMMARY

The purpose of the 1994 Jefferson Proving Ground demonstration was to evaluate the state of the art in UXO detection and remediation technology. For all airborne detection demonstrators, the detection capabilities have been statistically indistinguishable from zero.

EVALUATION OF OILTON 80-ACRE SITE

I. DEMONSTRATOR RESULTS

	R_{crit} = 2 m	R_{crit} = 5 m	R_{crit} = 10 m
Detection Capability			
overall detection ratio	0.02	0.08	0.21
ordnance detection ratio	0.03	0.07	0.18
nonordnance detection ratio	0	0.13	0.38
mistyped ordnance ratio	1.0	1.0	1.0
False Positive Ratio, *FPR*	*	*	*
False Negative Ratio, *FNR*	0	0	0
False Alarm Rate, *FAR*	1.93	1.90	1.83
Radial Accuracy, ΔR_{xy}	4.4	8.8	20.4
Depth Accuracy, ΔR_z	**	**	**

* Ratio is undefined (i.e., zero denominator).

** No depth information provided.

II. DISCUSSION OF DEMONSTRATOR PERFORMANCE

Oilton used a helicopter-mounted FLIR 2000 AB infrared imager. Infrared images were correlated to visual images simultaneously recorded by a CCD camera, which were then compared to surface landmarks. The entire 80-acre site was visited during the allotted time.

The detection capabilities calculated above are based on the probability of a one-to-one match between a demonstrator declaration and an emplaced target. This method of determining detection probability assumes a high degree of sensor resolution, which may not be justified. If a sensor cannot resolve signals resulting from two (or more) closely spaced objects, this approach penalizes the demonstrator for the lack of resolution. A more generous measure of detection capability would be the probability that an item will be found if holes of radius R_{crit} are dug on all demonstrator declarations, allowing for the possibility that multiple baseline targets may be found with a single demonstrator declaration. Since

neither of these measures exactly corresponds to the probability of detection in the conventional sense, we refer to them as P_{match} and P_{near}.

In the absence of detailed sensor models for the demonstrator, we use R_{crit} as a surrogate for resolution and devise a third measure. This third measure does not count as missed targets any baseline items which are unmatched by the one-to-one algorithm but are within R_{crit} of a demonstrator's hit. In other words, multiple baseline items within R_{crit} of a demonstrator declaration are aggregated into a single baseline item. This measure, P_{group}, is by construction intermediate between the other two, and the one most likely to prove a reasonable surrogate for detection as it is usually meant. The following table compares P_d calculated using these three measures, where P_{group} and P_{match} represent upper and lower bounds on the probability of detection achieved on this demonstration. The false alarms will be the same for P_{match}, P_{near}, and P_{group}. Also presented in the table are P_{match} and P_{near} calculated with the plastic mines removed from the baseline. All P_ds are for ordnance items only.

The detection probabilities were so low for the 80-acre site generally that it is important to correct the P_{match} computed by the target matching algorithm for the number of matches that would have been made by the same number of demonstrator declarations placed randomly on the site. The corrected P_{match} values are essentially zero or negative. Coupled with the small absolute number of matches, this shows that there is no statistically significant evidence that matching of this demonstrator's reported declarations to emplaced ordnance items is other than a random occurrence. The table is included in part to show how small detection capability is by any measure. Even when more generous measures such as P_{group} or P_{near} are considered, or when a very large R_{crit} is used to allow for inaccurate location ability, detection capabilities are insignificant.

P_d calculation method	R_{crit} = 2 m	R_{crit} = 5 m	R_{crit} = 10 m
P_{match}*	0.03	0.07	0.19
P_{match} (no mines)	0.03	0.08	0.21
P_{group}	0.03	0.07	0.21
P_{near}	0.03	0.09	0.30
P_{near} (no mines)	0.03	0.11	0.35

* For this demonstrator, corrections for random matches reduced P_{match} to 0.01 and – 0.09 for R_{crit} = 2 m and 5 m, respectively. For R_{crit} = 10 m, the approximations made in the correction become invalid, making the correction term artifically large.

To further support this notion, if matches are generated by randomly placed hits, the expected location accuracy is given by $\Delta R_{xy} = 2/3\ R_{crit}$. These calculated values, shown in the table below, match well to the observed location accuracies of this demonstrator, further supporting the notion that no demonstrator declarations are traceable to an emplaced ordnance item.

	R_{crit} = 2 m	R_{crit} = 5 m	R_{crit} = 10 m
ΔR_{xy} (ft)	4.4	8.8	20.4
2/3 R_{crit} (ft)	4.4	10.9	21.9

The standard deviation also scales with R_{crit} and has approximately the expected value.

Given that there is no confidence in the matches resulting from the emplaced ordnance, the classification capabilities for the airborne demonstrators were not computed.

The following table provides information about specific demonstrator declarations and baseline targets that may be of use to the demonstrator in assessing target identification capabilities for future operations, or for making improvements in the sensor or the data processing algorithms. The correct locations are provided for four mortars, three of which are on the surface. Three demonstrator declarations that were not within 5 m of a baseline item are also provided.

Locations and Types of Selected Baseline Targets				
Type	Size	X[a]	Y	Depth
Mortar	Small	2101548.4	14120689.8	2.06
Mortar	Small	2102121.0	14121733.6	0.0
Mortar	Small	2102061.1	14121777.9	0.0
Mortar	Small	2101976.8	14121862.1	0.0
[a] All dimensions are in feet. UTM Coordinates are U.S. Survey feet.				

Demonstrator Declarations Not Within 5 m of a Baseline Item	
72[a]	No known baseline item within 5 m of this location.
226	No known baseline item within 5 m of this location.
623	No known baseline item within 5 m of this location.
[a] Target numbers are from the demonstrator numbering scheme. Positions and types are correct for emplaced targets.	

III. SUMMARY

The purpose of the 1994 Jefferson Proving Ground demonstration was to evaluate the state of the art in UXO detection and remediation technology. For all airborne detection demonstrators, the detection capabilities have been statistically indistinguishable from zero. The Oilton demonstration had the highest false alarm rate by a significant amount. The relatively high P_{match} (compared with other airborne demonstrators) given in the summary report on the demonstration is a direct consequence of the large number of declarations.

EVALUATION OF SRI (FIXED WING) 80-ACRE SITE

I. DEMONSTRATOR RESULTS

	R_{crit} = 2 m	R_{crit} = 5 m	R_{crit} = 10 m
Detection Capability			
overall detection ratio	0.01	0.03	0.08
ordnance detection ratio	0.01	0.02	0.07
nonordnance detection ratio	0	0.08	0.17
mistyped ordnance ratio	0	0	0
False Positive Ratio, *FPR*	*	1.0	1.0
False Negative Ratio, *FNR*	0.98	0.97	0.90
False Alarm Rate, *FAR*	0.37	0.36	0.34
Radial Accuracy, ΔR_{xy}	4.6	10.3	20.7
Depth Accuracy, ΔR_z	8.8	5.1	4.0

* Ratio is undefined (i.e., zero denominator).

II. DISCUSSION OF DEMONSTRATOR PERFORMANCE

For this demonstration, SRI International used a ground penetrating radar mounted on a fixed wing aircraft. The radar returns were processed with a synthetic aperture algorithm to produce high-resolution images. Navigation was accomplished with an on-board global positioning receiver. The demonstrator visited the entire 80-acre site in the allotted time. The demonstrator report notes that the capabilities of the ground penetrating radar will be adversely affected by the wet ground conditions at JPG.

The detection capabilities calculated above are based on the probability of a one-to-one match between a demonstrator declaration and an emplaced target. This method of determining detection probability assumes a high degree of sensor resolution, which may not be justified. If a sensor cannot resolve signals resulting from two (or more) closely spaced objects, this approach penalizes the demonstrator for the lack of resolution. A more generous measure of detection capability would be the probability that an item will be found if holes of radius R_{crit} are dug on all demonstrator declarations, allowing for the possibility

that multiple baseline targets may be found with a single demonstrator declaration. Since neither of these measures exactly corresponds to the probability of detection in the conventional sense, we refer to them as P_{match} and P_{near}.

In the absence of detailed sensor models for the demonstrator, we use R_{crit} as a surrogate for resolution and devise a third measure. This third measure does not count as missed targets any baseline items which are unmatched by the one-to-one algorithm but are within R_{crit} of a demonstrator's hit. In other words, multiple baseline items within R_{crit} of a demonstrator declaration are aggregated into a single baseline item. This measure, P_{group}, is by construction intermediate between the other two, and the one most likely to prove a reasonable surrogate for detection as it is usually meant. The following table compares P_d calculated using these three measures, where P_{group} and P_{match} represent upper and lower bounds on the probability of detection achieved on this demonstration. The false alarms will be the same for P_{match}, P_{near}, and P_{group}. Also presented in the table are P_{match} and P_{near} calculated with the plastic mines removed from the baseline. All P_ds are for ordnance items only.

The detection probabilities were so low for the 80-acre site generally that it is important to correct the P_{match} computed by the target matching algorithm for the number of matches that would have been made by the same number of demonstrator declarations placed randomly on the site. The corrected P_{match} values are essentially zero or negative. Coupled with the small absolute number of matches, this shows that there is no statistically significant evidence that matching of this demonstrator's reported declarations to emplaced ordnance items is other than a random occurrence. The table is included in part to show how small detection capability is by any measure. Even when more generous measures such as P_{group} or P_{near} are considered, or when a very large R_{crit} is used to allow for inaccurate location ability, detection capabilities are insignificant.

P_d calculation method	R_{crit} = 2 m	R_{crit} = 5 m	R_{crit} = 10 m
P_{match}*	0.01	0.02	0.07
P_{match} (no mines)	0.01	0.03	0.08
P_{group}	0.01	0.02	0.08
P_{near}	0.01	0.04	0.20
P_{near} (no mines)	0.01	0.05	0.23

* For this demonstrator, corrections for random matches reduced P_{match} to 0.01 and – 0.01 for R_{crit} = 2 m and 5 m, respectively. For R_{crit} = 10 m, the approximations made in the correction become invalid, making the correction term artifically large.

To further support this notion, if matches are generated by randomly placed hits, the expected location accuracy is given by $\Delta R_{xy} = 2/3\ R_{crit}$. These calculated values, shown in the table below, match well to the observed location accuracies of this demonstrator, further supporting the notion that no demonstrator declarations are traceable to an emplaced ordnance item.

	$R_{crit} = 2$ m	$R_{crit} = 5$ m	$R_{crit} = 10$ m
ΔR_{xy} *(ft)*	4.6	10.3	20.7
2/3 R_{crit} *(ft)*	4.4	10.9	21.9

The standard deviation also scales with R_{crit} and has approximately the expected value.

Given that there is no confidence in the matches resulting from the emplaced ordnance, the classification capabilities for the airborne demonstrators were not computed.

The following table provides information about specific demonstrator declarations and baseline targets that may be of use to the demonstrator in assessing target identification capabilities for future operations, or for making improvements in the sensor or the data processing algorithms. The correct locations are provided for four mortars, three of which are on the surface. Three demonstrator declarations that were not within 5 m of a baseline item are also provided.

Locations and Types of Selected Baseline Targets				
Type	Size	X[a]	Y	Depth
Mortar	Small	2101548.4	14120689.8	2.06
Mortar	Small	2102121.0	14121733.6	0.0
Mortar	Small	2102061.1	14121777.9	0.0
Mortar	Small	2101976.8	14121862.1	0.0
[a] All dimensions are in feet. UTM Coordinates are U.S. Survey feet.				

Demonstrator Declarations Not Within 5 m of a Baseline Item	
34[a]	No known baseline item within 5 m of this location.
75	No known baseline item within 5 m of this location.
134	No known baseline item within 5 m of this location.
[a] Target numbers are from the demonstrator numbering scheme. Positions and types are correct for emplaced targets.	

III. SUMMARY

The purpose of the 1994 Jefferson Proving Ground demonstration was to evaluate the state of the art in UXO detection and remediation technology. For all airborne detection demonstrators, the detection capabilities have been statistically indistinguishable from zero.

EVALUATION OF SRI (ROTARY WING) 80-ACRE SITE

I. DEMONSTRATOR RESULTS

	$R_{crit} = 2$ m	$R_{crit} = 5$ m	$R_{crit} = 10$ m
Detection Capability			
overall detection ratio	0.005	0.04	0.06
ordnance detection ratio	0.006	0.03	0.04
nonordnance detection ratio	0	0.13	0.16
mistyped ordnance ratio	0	0	0
False Positive Ratio, *FPR*	*	1.0	1.0
False Negative Ratio, *FNR*	0.99	0.95	0.91
False Alarm Rate, *FAR*	0.28	0.28	0.26
Radial Accuracy, ΔR_{xy}	3.0	11.1	15.4
Depth Accuracy, ΔR_z	**	**	**

* Ratio is undefined (i.e., zero denominator).

** No depth information provided.

II. DISCUSSION OF DEMONSTRATOR PERFORMANCE

For this demonstration, SRI, international used an ultra wideband, bistatic ground penetrating radar on a rotary wing platform. The sensor performance, which has maximum depth of 10 meters in dry sand, was limited by the wet, clay soil conditions at JPG. The global positioning system was used for navigation. The entire 80-acre site was visited during the allotted time.

The detection capabilities calculated above are based on the probability of a one-to-one match between a demonstrator declaration and an emplaced target. This method of determining detection probability assumes a high degree of sensor resolution, which may not be justified. If a sensor cannot resolve signals resulting from two (or more) closely spaced objects, this approach penalizes the demonstrator for the lack of resolution. A more generous measure of detection capability would be the probability that an item will be found if holes of radius R_{crit} are dug on all demonstrator declarations, allowing for the possibility

that multiple baseline targets may be found with a single demonstrator declaration. Since neither of these measures exactly corresponds to the probability of detection in the conventional sense, we refer to them as P_{match} and P_{near}.

In the absence of detailed sensor models for the demonstrator, we use R_{crit} as a surrogate for resolution and devise a third measure. This third measure does not count as missed targets any baseline items which are unmatched by the one-to-one algorithm but are within R_{crit} of a demonstrator's hit. In other words, multiple baseline items within R_{crit} of a demonstrator declaration are aggregated into a single baseline item. This measure, P_{group}, is by construction intermediate between the other two, and the one most likely to prove a reasonable surrogate for detection as it is usually meant. The following table compares P_d calculated using these three measures, where P_{group} and P_{match} represent upper and lower bounds on the probability of detection achieved on this demonstration. The false alarms will be the same for P_{match}, P_{near}, and P_{group}. Also presented in the table are P_{match} and P_{near} calculated with the plastic mines removed from the baseline. All P_ds are for ordnance items only.

The detection probabilities were so low for the 80-acre site generally that it is important to correct the P_{match} computed by the target matching algorithm for the number of matches that would have been made by the same number of demonstrator declarations placed randomly on the site. The corrected P_{match} values are essentially zero or negative. Coupled with the small absolute number of matches, this shows that there is no statistically significant evidence that matching of this demonstrator's reported declarations to emplaced ordnance items is other than a random occurrence. The table is included in part to show how small detection capability is by any measure. Even when more generous measures such as P_{group} or P_{near} are considered, or when a very large R_{crit} is used to allow for inaccurate location ability, detection capabilities are insignificant.

P_d calculation method	R_{crit} = 2 m	R_{crit} = 5 m	R_{crit} = 10 m
P_{match}*	0.01	0.03	0.04
P_{match} (no mines)	0.01	0.03	0.05
P_{group}	0.01	0.03	0.05
P_{near}	0.01	0.07	0.19
P_{near} (no mines)	0.01	0.08	0.22

* For this demonstrator, corrections for random matches reduced P_{match} to 0.0 and 0.01 for R_{crit} = 2 m and 5 m, respectively. For R_{crit} = 10 m, the approximations made in the correction become invalid, making the correction term artifically large.

If matches are generated by randomly placed hits, the expected location accuracy is given by $\Delta R_{xy} = 2/3\ R_{crit}$. These calculated values, shown in the table below, match well to the observed location accuracies of this demonstrator, further supporting the notion that no demonstrator declarations are traceable to an emplaced ordnance item.

	R_{crit} = 2 m	R_{crit} = 5 m	R_{crit} = 10 m
ΔR_{xy} (ft)	3.0	11.1	15.4
2/3 R_{crit} (ft)	4.4	10.9	21.9

The standard deviation also scales with R_{crit} and has approximately the expected value.

Given that there is no confidence in the matches resulting from the emplaced ordnance, the classification capabilities for the airborne demonstrators were not computed.

The following table provides information about specific demonstrator declarations and baseline targets that may be of use to the demonstrator in assessing target identification capabilities for future operations, or for making improvements in the sensor or the data processing algorithms. The correct locations are provided for four mortars, three of which are on the surface. Three demonstrator declarations that were not within 5 m of a baseline item are also provided.

Locations and Types of Selected Baseline Targets				
Type	Size	X[a]	Y	Depth
Mortar	Small	2101548.4	14120689.8	2.06
Mortar	Small	2102121.0	14121733.6	0.0
Mortar	Small	2102061.1	14121777.9	0.0
Mortar	Small	2101976.8	14121862.1	0.0
a All dimensions are in feet. UTM Coordinates are U.S. Survey feet.				

Demonstrator Declarations Not Within 5 m of a Baseline Item	
41[a]	No known baseline item within 5 m of this location.
80	No known baseline item within 5 m of this location.
107	No known baseline item within 5 m of this location.
a Target numbers are from the demonstrator numbering scheme. Positions and types are correct for emplaced targets.	

III. SUMMARY

The purpose of the 1994 Jefferson Proving Ground demonstration was to evaluate the state of the art in UXO detection and remediation technology. For all airborne detection demonstrators, the detection capabilities have been statistically indistinguishable from zero.

7. REMEDIATION EVALUATIONS

This page is intentionally left blank

EVALUATION OF BENTHOS, INC.
REMEDIATION

I. DEMONSTRATOR RESULTS

The results of target excavation for this demonstrator are shown in the following table.

Target Name	Class	Depth (feet)	Success[1]	Volume[2] (yds^3)	Duration[3] (hrs)	Remote Hazard Assessment[4]
A-0029 60 mm	Mortar	0.89	Y	1.00	1.00	N
H-0027 106 mm	Bomb	3.75	Y	1.50	2.00	NA
P-0002 250#	Bomb	2.94	Y	1.00	2.00	Y
G-0003 155 mm	Projectile	2.70	Y	1.00	1.50	Y
A-0046 60 mm	Projectile	3.10	Y	1.75	2.00	NA
E-0013 90 mm	Projectile	6.10	Y	2.50	3.50	Y
A-0036 60 mm[5]	Mortar	0.79	Y	0.10	10.00	Y
B-0034 81 mm	Projectile	1.23	Y	0.75	0.30	Y
I-0011 175 mm	Projectile	1.99	Y	1.00	0.40	Y
F-0026 152 mm	Projectile	3.00	Y	1.00	1.00	Y
P-0007 250#	Bomb	5.60	Y	2.00	2.30	Y

1 Y= successfully remediated. N = not successfully remediated. P = partial success.

2 The volume of earth, in cubic yards, that was excavated in remediating the target.

3 Estimate of the time for remediating the target.

4 Indicates whether the demonstrator examined the target remotely, prior to moving or directly contacting the ordnance item. Y = remotely assessed. N = not remotely assessed. NA = information not provided by the demonstrator.

5 The reported values for volume of earth excavated and duration of operation for this ordnance item are grossly different from those reported for target A-0029, which was the same target class (mortar) at a similar depth.

II. DISCUSSION OF DEMONSTRATOR PERFORMANCE

Benthos, Inc. used a Remote Dig-It system for remediation of targets. The Dig-It system is a remotely operated backhoe excavator that is controlled using a hand-held controller connected to the backhoe via a 25-foot cable. The excavator has three on-board

cameras that convey visual information back to video monitors in the control system. The excavator navigated using a differential global positioning system.

Benthos was assigned 12 targets to excavate. As indicated in above, 11 targets were successfully remediated. The twelfth location contained no ordnance items. Eight of the ordnance items were examined by the remote excavator for hazard assessment. One target was first identified during the dumping of soil and for the remaining two, no comments were made regarding the state of the target.

III. SUMMARY

Benthos, Inc. successfully remediated the 12 assigned targets in 28 of the allotted 40 hours. Eleven locations contained ordnance items and one did not. Of the 11 ordnance targets, eight were examined by the remote excavator, prior to direct handling.

EVALUATION OF SANDIA NATIONAL LABORATORIES REMEDIATION

I. DEMONSTRATOR RESULTS

The results of target excavation for this demonstrator are shown in the following table.

Target Name	Class	Depth (feet)	Success[1]	Volume[2] (yds^3)	Duration[3] (hrs)	Remote Hazard Assessment[4]
488 AP Mine	Mine	0.2	Y	0.07	24.00	NA
489 AP Mine	Mine	0.2	Y	0.02	2.00	NA
490 AP Mine	Mine	0.2	Y	0.02	2.00	NA
491 AP Mine	Mine	0.2	Y	0.02	2.00	NA
492 AP Mine	Mine	0.2	Y	0.02	2.00	NA
414 unknown	Mortar	3	P	0.07	2.00	NA
477 unknown	Other	2	P	0.33	60.00	NA
420 unknown	Other		P			NA

1 Y= successfully remediated. N = not successfully remediated. P = partial success.

2 The volume of earth, in cubic yards, that was excavated in remediating the target.

3 Estimate of the time for remediating the target.

4 Indicates whether the demonstrator examined the target remotely, prior to moving or directly contacting the ordnance item. Y = remotely assessed. N = not remotely assessed. NA = information not provided by the demonstrator.

II. DISCUSSION OF DEMONSTRATOR PERFORMANCE

Sandia National Laboratories used a Remote Telerobotic Vehicle for Intelligent Remediation (RETVIR) of targets. The vehicle has two heads that can be attached to a pincer arm: a shovel head for excavating and a sensor head for measuring magnetic fields. The excavator is remotely operated from a control station that contains a hand controller, video monitors, and status indicators. The excavator navigated using a differential global positioning system. Excavator performance is limited by hard soil, rainy weather, and rough terrain. Equipment failures caused over 7 hours of down-time. Radio interference between the data links of the Wright Labs and Sandia Labs systems forced the Sandia excavator to move to the 80-acre site.

Sandia was assigned to excavate 8 targets. Five of the targets were successfully remediated. These targets were antipersonnel mines buried less than 1 foot below the surface. The projectile, cluster, and mortar targets assigned were not successfully remediated. The projectile was not found and the other two targets were buried too deep for the capabilities of the excavator. The times reported by this demonstrator vary widely. The reported 24-hour and 60-hour times are not excavation times, but rather are elapsed times. These items were not located on the first try; the second try, which was successful, occurred 24 hours and 60 hours later.

III. SUMMARY

Sandia successfully remediated five of the assigned eight targets in the allotted 40 hours.

EVALUATION OF WRIGHT LABORATORIES REMEDIATION

I. DEMONSTRATOR RESULTS

The results of target excavation for this demonstrator are shown in the following table.

Target Name	Class	Depth (feet)	Success[1]	Volume[2] (yds^3)	Duration[3] (hrs)	Remote Hazard Assessment[4]
Unknown anomaly	Other		N	NA		NA
Unknown anomaly	Other	1.10	Y	NA	1.03	NA
Unknown anomaly	Other		Y	NA	0.30	NA
Unknown anomaly	Other		N	NA		NA
B-0007 81 mm	Mortar	1.14	Y	NA	1.29	NA
K-0018 500 lb	Bomb	7.95	N	NA	2.14	NA
A-0033 60 mm	Mortar	0.94	Y	NA	1.12	NA
F-0029	Mortar	5.14	N	NA		NA
B-0043 81 mm	Mortar	0.92	N	NA		NA

1 Y= successfully remediated. N = not successfully remediated. P = partial success.

2 The volume of earth, in cubic yards, that was excavated in remediating the target. This information was not provided by this demonstrator.

3 Estimate of the time for remediating the target.

4 Indicates whether the demonstrator examined the target remotely, prior to moving or directly contacting the ordnance item. Y = remotely assessed. N = not remotely assessed. NA = information not provided by the demonstrator.

II. DISCUSSION OF DEMONSTRATOR PERFORMANCE

Wright Laboratories used a remotely operated Caterpillar 325 long reach excavator fitted with a 3-foot general purpose bucket and clam thumb for remediation of targets. The excavator navigated using a differential global positioning system. The excavator has two

on-board cameras that convey visual information back to video monitors in the control system. Equipment problems caused some periods of down-time.

Wright Laboratory was assigned 9 targets to excavate. Four of the targets were successfully remediated. Four of the targets were not excavated and the demonstrator failed to located one target. The demonstrator did not report remote assessments.

III. SUMMARY

Wright Labs successfully remediated four of the nine assigned targets in 3.74 hours. Another 2.14 hours were spent excavating a fifth target, which the demonstrator was not successful at remediating. An unknown amount of time was spent investigating the remaining four targets.

8. DEMONSTRATION COSTS AND SURVEY RATES

As part of the proposal submitted for the JPG demonstration, demonstrators were required to submit firm fixed-price quotations to the government. These quotations were to include the cost of the transportation, personnel, and operations on the JPG sites. In many cases, the quoted prices are dominated by travel costs. The costs were to reflect the price of surveying the entire site assigned. Because many of the demonstrators searched only a portion of their assigned site, the quoted costs are paired with the area searched (in acres).

These cost quotations should not be taken as a true representation of the costs of conducting a clean-up or even a sweep of a comparable site. The demonstrators were permitted only to survey the site. They were not permitted to pick up debris prior to operations, nor were they permitted to sweep the site to remove objects detected after one pass so that follow-on passes could be conducted with higher sensitivity.

Despite the above considerations, the quotations may be an indication of the cost of doing surveys similar to those done at JPG to support decisions about whether site remediation is warranted. Finally, the cost figures do not even attempt to include the cost of remediating targets once they are detected, which will dominate the cost of any site cleanup.

Demonstrator	Cost ($)	Area Searched (acres)
40-Acre Site		
Areté Engineering Technology	146,500	25
Australian Defence Industries	87,580	40
Battelle and OSU	85,189	2.3 [b]
Chemrad/EG&G	157,001	16
Chemrad (G-822L)	46,384	40
Chemrad (GSM-19)	45,085	40
Coleman Research Corp.	62,111	36
Dynamic Systems, Inc.	74,234	5.5
ENSCO	98,480	10
EODT	39,227	11 [a]
Foerster	247,826	24
GDE Systems	43,042	6.9
Geo Centers	151,949	40
GeoMetrics, Inc.	70,438	35
GeoRadar, Inc.	21,398	2
Jaycor	96,198	20 [a]
Metratek, Inc.	49,250	5
Security Search Products (Vallon)	56,750	12
SRI, Surface-Towed	126,577	13
UXB International	32,960	30 [a]
80-Acre Site		
Airborne Environmental	126,474	80
Battelle	85,189	29
Geonex Aerodat Inc.	22,156	80
Oilton	165,222	80
SRI (rotary wing)	81,021	80
SRI (fixed wing)	143,389	80

a Proposal indicated that the demonstrator would be unable to visit the entire site within the allotted 1 week.

b Proposal did not indicate the expected survey rate.

9. REFERENCES

1. Demonstrator Work Plan for the UXO Detection, Identification, and Remediation Advanced Technology Demonstration at the Jefferson Proving Ground, Indiana, PRC, Inc., NAVEODTECHCEN Contract No. N00600-88-D-3717, Delivery Order FG-3S.

2. *Unexploded Ordnance Advanced Technology Demonstration Program at Jefferson Proving Ground (Phase 1)*, PRC, Inc., AEC Report No. SFIM-AEC-ET-CR-94120.

3. *Assessment of Demonstrator Performance at the Jefferson Proving Ground Unexploded Ordnance Demonstration*, Institute for Defense Analyses, IDA Document D-1687, to be published.

4. Van Trees, H.L., *Detection, Estimation, and Modulation Theory*, Part I, 1968, John Wiley & Sons, Inc.

This page is intentionally left blank